

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 09:34:42 ; Search time 285 Seconds

(without alignments)
55.302 Million cell updates/sec

Title: US-10-020-674-2

Perfect score: 2013
Sequence: 1 MKGLLYGTNDIRSETYPE.....QLIEHKNNVKKIIVTPEVS 380

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSPROT.40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	923.5	45.9	382	1	BDH_YEAST	P39714 saccharomyc
2	912	45.3	417	1	YAG1_YEAST	P39713 saccharomyc
3	408.5	20.3	352	1	DHSD_BACSU	Q06004 bacillus su
4	370.5	18.4	348	1	DHSD_BOMMO	Q02912 bombix mori
5	362	18.0	360	1	DHSD_SCHPO	P36524 schizosacch
6	360.5	17.9	357	1	DHSD_YEAST	P35697 saccharomyc
7	342	17.0	358	1	YDUL_ECOLI	P77539 escherichia
8	342	17.0	363	1	XYL2_PICST	P22144 pichia stip
9	338	16.8	356	1	DHSD_HUMAN	Q00796 homo sapien
10	337.5	16.8	343	1	DHSD_BACHD	Q92901 bacillus ha
11	337	16.7	347	1	YDUL_ECOLI	P77280 escherichia
12	336.5	16.7	354	1	DHSD_SHEEP	P07846 ovis aries
13	331.5	16.5	399	1	DHSD_RAT	P27667 rattus norv
14	329	16.3	375	1	DHSD_MOUSE	Q64442 mus musculu
15	319.5	15.9	339	1	RSPB_ECOLI	P38105 escherichia
16	316	15.7	337	1	YDUL_ECOLI	P39400 escherichia
17	310	15.4	344	1	TDH_RHIME	Q52998 ribicobium m
18	307.5	15.3	346	1	GATD_ECOLI	P37190 escherichia
19	306	15.2	342	1	Y053_HAETN	Q57517 haemophilus
20	306	15.2	424	1	FADH_METWR	P47734 methyllobact
21	297	14.8	347	1	TDH_BACSU	Q31776 bacillus str
22	279.5	13.9	340	1	TDH_XANCP	Q34268 xanthomonas
23	271	13.5	412	1	YDUL_ECOLI	P77316 escherichia
24	269	13.4	341	1	TDH_ECOLI	P07913 escherichia
25	265.5	13.2	343	1	TDND_ECOLI	P39346 escherichia
26	261	13.0	379	1	FADH_PICPA	Q74685 pichia past
27	257	12.8	339	1	ADH3_BACST	P43228 bacillus st
28	257	12.8	369	1	ADH3_SYNY3	P73138 synecchocyst
29	255.5	12.7	375	1	ADH2_PERMA	P41681 peromyscus
30	255	12.7	369	1	ADH3_ECOLI	P25437 escherichia
31	253.5	12.6	360	1	FADH_AMYME	P80092 amycolatops
32	250	12.4	359	1	MTD2_ARATH	Q02972 arabidopsisi
33	248	12.3	337	1	ADH1_BACST	P12311 bacillus str

34	246	12.2	336	1	ADHP_ECOLI	P39451	eschlerichia
35	246	12.2	337	1	ADH1_ZYMMO	P20368	zymomonas m
36	246	12.2	339	1	ADH2_BACST	P42327	bacillus st
37	245	12.2	357	1	MTD1_ARATH	O02971	arabidopsis
38	245	12.2	379	1	ADH2_STPCA	P80468	struthio ca
39	244	12.1	378	1	ADH3_HAETN	P44557	haemophilus
40	243	12.1	376	1	ADH3_SPAUV	P73866	spatius aura
41	242.5	12.0	353	1	YPHC_ECOLI	P7360	eschlerichia
42	241.5	12.0	365	1	FADH_YEAST	P32771	saccharomyc
43	239.5	11.9	366	1	ADH_ALCEU	P14940	alcaligenes
44	239	11.9	378	1	ADH3_DROME	P46415	drosophila
45	238.5	11.8	350	1	ADH_SCHRO	P00332	schizosacch

ALIGNMENTS

RESULT 1
BDH_YEAST STANDARD: PRT: 382 AA.
ID BDH_YEAST
AC P39714;
DT 01-FEB-1995 (Rel. 31, Last Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE (R,R)-butanediol dehydrogenase (EC 1.1.1.4).
GN BDH OR YAL060W OR FUN49.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288c / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kackabek D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
cerevisiae";
RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RL [2]
RN CHARACTERIZATION:
RP STRAIN=FY834a1pha;
RX MEDLINE=20549593; PubMed=10938079;
RA Gonzalez E., Fernandez M.R., Larroy C., Sola L., Pelicas M.A.,
RA Pares X., Biosca J.A.;
RT "Characterization of a (2R,3R)-2,3-butanediol dehydrogenase as the
Saccharomyces cerevisiae YAL060W gene product. Disruption and
induction of the gene";
RT J. Biol. Chem. 275:35876-35885(2000).
RL -1- CATALYTIC ACTIVITY: (R,R)-butane-2,3-diol + NAD(+) = (R)-acetoin +
NADH.
CC -1- COFACTOR: Zinc (By similarity).
CC -1- SUBUNIT: Homodimer.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U12880; AAC04974.1; -;
CC SGD: S0000056; BDH.
CC InterPro: IPR002328; ADH_zinc.
CC InterPro: IPR02085; Adh_zn_family.
CC Pfam: PF00107; adh_zinc; 1.
CC PROSITE: PS00059; ADH_ZINC; 1.
CC Oxioreductase; Zinc; NAD.
CC METAL 39 39 ZINC 1 (BY SIMILARITY).
FT

FT METAL 73 73 ZINC 1 (BY SIMILARITY).
 FT METAL 103 103 ZINC 2 (BY SIMILARITY).
 FT METAL 120 120 ZINC 2 (BY SIMILARITY).
 FT METAL 123 123 ZINC 2 (BY SIMILARITY).
 FT METAL 131 131 ZINC 2 (BY SIMILARITY).
 FT METAL 173 173 ZINC 1 (BY SIMILARITY).
 SQ SEQUENCE 382 AA; 41538 MW; 1006FA596DE91A17 CRC64;

Query Match 45.9%; Score 923.5; DB 1; Length 382;
 Best Local Similarity 48.7%; Pred. No. 1,1e-63;
 Matches 184; Conservative 65; Mismatches 126; Indels 3; Gaps 2;

QY 1 MKGLYYGTNDIRYSETPPEIKNPNDYKIKSYCGICGTDLKEFTYSGGPVFPKQGT 60
 1 MRALAFKKGDIHFTNDIPREIOTDEVIIDVSCGICGSDLHE--YLDGPIFMKDE 58
 DB 179 WHAVKISGKRGSSALVLGAGPIGLCTIYLKMGASKIYVSEIARRIEMAKKLEVEF 238
 QY 61 KDKISGTEPLPCGHEFSGTVEVSGVTSVKPGRVAVEATSHCSDRSRKYDTVAQDLG 120
 59 CHRLSNALPLAMGHSMGSGIVSKYKPKVKVGDHVVYDAASCADLHCMPHSKRYNSK 118
 QY 121 LCMACSGSPNCASLSPFGILGASGGAFAEYVYGGEDHWKLPDSTIPDDIGALVEPISA 180
 119 PCACACRGSENLCTHAGFTGLVTSGFAEQVYVSOHIIIVKEIPLDVAALVEPLSVT 178
 QY 181 WHAVERARPOQTALVYLGGPIGLATITLALOGHNAKIVCSEPALIRROFAKEIGAEVF 240
 179 WHAVKISGKRGSSALVLGAGPIGLCTIYLKMGASKIYVSEIARRIEMAKKLEVEF 238
 QY 241 DPSTCDAN-AVLKAVPENEGFAHAFDCSGVPQTFTTSIVATGPGIANNVAVMGDHP 299
 239 NSKHGHSKSTELRLGLTKSHDGFYSIDSGIQTVEFETSLKALTFKGTATNIAVWPKPV 298
 QY 300 GFMPSLTYOEKRYATGSCYTVKDFQEVYKALDEGLISLDRKAMITGKVLKDGVEKG 359
 299 PQGPMVDTLOEKWMSIGYVFAFEVYRAIHGDIAMEDCQILITGKORIDGWEKG 358
 DB 360 KOLIEKKNVKTLPN 377
 359 OELMDKESNVKILTPN 376

RESULT 2
 YAG1_YEAST STANDARD; PRT; 417 AA.

AC P39713;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Hypothetical zinc-type alcohol dehydrogenase-like protein in GDH3-CNE1
 GN Intergenic region.
 GN YAL061W OR F0N50.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=95249563; PubMed=7731988;
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
 RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
 RA Storms R.K.;
 RT "The nucleotide sequence of chromosome I from Saccharomyces
 RT cerevisiae";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 RL -1- CORACTOR: ZINC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U12980; AAC04973.1; -
 DR SGD: S0000057; YAL061W.
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zn_family.
 DR Pfam: PF00107; adh_zinc; 1.
 DR PROSITE: PS00059; ADH_ZINC; 1.
 KW Hypothetical protein; Oxidoreductase; zinc.
 FT METAL 39 39 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 64 64 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 120 120 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 123 123 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 131 131 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 173 173 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 417 AA; 46098 MW; D921CEBDE61D0151 CRC64;

Query Match 45.3%; Score 912; DB 1; Length 417;
 Best Local Similarity 45.9%; Pred. No. 9.2e-63;
 Matches 174; Conservative 71; Mismatches 130; Indels 4; Gaps 2;

QY 1 MKGLYYGTNDIRYSETPPEIKNPNDYKIKSYCGICGTDLKEFTYSGGPVFPKQGT 60
 1 MRALAFKKGDIHFTNDIPREIOTDEVIIDVSCGICGSDLHE--YLDGPIFMKDE 58
 DB 179 WHAVKISGKRGSSALVLGAGPIGLCTIYLKMGASKIYVSEIARRIEMAKKLEVEF 238
 QY 61 KDKISGTEPLPCGHEFSGTVEVSGVTSVKPGRVAVEATSHCSDRSRKYDTVAQDLG 120
 59 CHRLSNALPLAMGHSMGSGIVSKYKPKVKVGDHVVYDAASCADLHCMPHSKRYNSK 118
 QY 121 LCMACSGSPNCASLSPFGILGASGGAFAEYVYGGEDHWKLPDSTIPDDIGALVEPISA 180
 119 PCACACRGSENLCTHAGFTGLVTSGFAEQVYVSOHIIIVKEIPLDVAALVEPLSVT 178
 QY 181 WHAVERARPOQTALVYLGGPIGLATITLALOGHNAKIVCSEPALIRROFAKEIGAEVF 240
 179 WHAVKISGKRGSSALVLGAGPIGLCTIYLKMGASKIYVSEIARRIEMAKKLEVEF 238
 QY 241 DPS--TCDDANAVLKAMVNEGFAHAFDCSGVPQTFTTSIVATGPGIANNVAVMGDHP 298
 239 NSKHGHSKSTELRLGLTKSHDGFYSIDSGIQTVEFETSLKALTFKGTATNIAVWPKPV 298
 QY 299 GFMPSLTYOEKRYATGSCYTVKDFQEVYKALDEGLISLDRKAMITGKVLKDGVEKG 358
 299 PQGPMVDTLOEKWMSIGYVFAFEVYRAIHGDIAMEDCQILITGKORIDGWEKG 358
 DB 360 KOLIEKKNVKTLPN 377
 359 IMKLINKESTIKILTPN 377

RESULT 3

DHSO_BACSU STANDARD; PRT; 352 AA.
 AC Q06004;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorbitol dehydrogenase (EC 1.1.1.14) (L-iditol 2-dehydrogenase)
 GN (Glucitol dehydrogenase).
 GN GUPB.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-31.
 RC STRAIN=168;
 RX MEDLINE=93094198; PubMed=1460002;
 RA Ng K., Ye R., Wu X.-C., Wong S.-L.;
 RA "Sorbitol dehydrogenase from Bacillus subtilis. Purification,
 RA characterization, and gene cloning";
 RT J. Biol. Chem. 267:24989-24994(1992).

Db 114 MEFAATPPYDGLTLYTITTEDPCTKLPRKQISVEGALFEPMASVAHVAMTRGNLKCGRV 173
 QY 196 LVYGGPGLATILALOGHHAGKIVCGSEPALIRBOPA-KEGAEVDPD-----STCDD 247
 Db 174 LVMGCGIVGLMMVAAYGAIDIVADSPSRVFEPAKIVGAPFPPIAKENESLPDY 233
 QY 248 ANAALKAMVENESEGFHAFAFCGVPQTFSTTSIVATGPGSIANVAWGDHPIGFPMPSLT 307
 Db 234 AQRKQALIEKGEFEDFADVATGICHTAVLNLKRGTFVQ-AGNCKPYIDFPINII 292
 QY 308 YOEKATGSMCTYKDFOEYVKALEDGLISLDRKAMITGVHLKDGVEKFKOLIEHKE 367
 Db 293 NVEINLVGSFRYAHGCKYKSLFVNSGLVDY---KPLITHFAFKDAL-KAYETVASGEE 348
 QY 368 NNKILV 374
 Db 349 GVLKVTI 355

RESULT 6

DH5O_YEAST STANDARD; PRT: 357 AA.
 ID P35497;

DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Sorbitol dehydrogenase (EC 1.1.1.14) (L-Iditol 2-dehydrogenase).
 GN SORL OR SDH1 OR YJR159W OR J2395.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94171068; PubMed=8125328.
 RA Searby A.V., Schopp C., Idler K.B.;
 RT Cloning and sequence determination of the gene encoding sorbitol
 RT dehydrogenase from *Saccharomyces cerevisiae*.;
 RL Gene 140:121-126(1994).
 RN [2]
 RP SEQUENCE OF 1-20 FROM N.A.
 RA Huang M.-E., Chuat J.-C., Galibert F.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 11-357 FROM N.A.
 RA Medler H., Underwood A.P., Louis E.J., Wambutt R.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-Iditol + NAD(+) -> L-sorbose + NADH.
 CC -1- COFACTOR: ZINC.
 CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: L11039; AAA35027.1; -;
 DR EMBL: Z49659; CAA89692.1; -;
 DR HSSP: P07846; ISDG.
 DR SGD: S0003920; SORL.
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zn_family.
 DR InterPro: IPR000205; NAD_binding.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_ZINC.1.
 FT Oxioreductase; zinc; NAD.
 FT METAL 43
 FT METAL 68 68 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 357 AA; 38165 MW; 16843C2040A334A CRC64;

Query Match 17.98; Score 360.5; DB 1; Length 357;
 Best Local Similarity 30.38; Pred. No. 1.7e-20;
 Matches 11; Conservative 56; Mismatches 160; Indels 39; Gaps 11;

QY 16 ETVEPEELKNNDYKIKVSYCGICGTDLKEFTYSQGPVFPFKQTKDKISGVEL--PLCP 73
 Db 20 EQRPIPTIKDPHYVLAIKATGICGSDI-HYRSGG-----IGKTIKAPVL 66
 QY 74 GHEPSGTVEVSGSVTSKPGDRVAVENTSHCSDRSKYDTVAODGLCMACQSGSPNC 133
 Db 67 GHESGQVVEYGDVATRVKVDRAVEP---GVPSRSDEN-----KEGRYMLC 112
 QY 134 ASLSFGGLGASGSGFAEYVYGEDHMKVLPSPIDPDIALYEPISVAMHAEARAFQCG 193
 Db 113 PHMAFAATPPIDGLVLYKYLSPEDLVKLPGVSYEGACVPEPLSVGHSKLAGVRGT 172
 QY 194 TALVYGGPGLATILALOGHHAGKIVCGSEPALIRBOPAKELGA-EVDPDS--TCDDANA 250
 Db 173 KVVYFGAGPVGLITGAVARAFCATDIVDVDFDNKLRKAKDFGATNTNSSQFSDKAD 232
 QY 251 VLKAMVENESEFHA-AEDCGVPQTFSTTSIVATGPGSIANVAWGDHPIGFPMPSLT 308
 Db 233 LADGVOKLLGNHADVVEGSGADVCDIAAVKTKVSGTMOVGM-GKNTNFPPIAEVSG 291
 QY 309 QEKATGSMCTYVDFOGVVKALEDGLISLDRKAMITGVHLKDGVEKFKOLIEHKE 368
 Db 292 KEMKLIGFRYSFGDRDAVNLVATGVNV--KPLITHKRFEDAA-KAYDYNIAHGE 347
 QY 369 NNKILV 374
 Db 348 VVKITII 353

RESULT 7

YDJL_ECOLI STANDARD; PRT: 358 AA.
 ID YDJL_ECOLI

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical zinc-type alcohol dehydrogenase-like protein ydjL.
 GN YDJL OR B1776.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley W., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
 RA Saito N., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.,
 RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -1- COFACTOR: ZINC (Potential).

CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC
DR EMBL: AE000272; AAC74846.1; -
DR EMBL: D90821; BAAL574.1; -
DR Ecogene: EG13488; ydJL.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Hypothetical protein; Oxidoreductase; Zinc; Complete proteome.
FT METAL 39
FT METAL 39
FT METAL 61
FT METAL 92
FT METAL 92
FT METAL 95
FT METAL 95
FT METAL 98
FT METAL 106
FT METAL 106
FT METAL 157
SQ SEQUENCE 358 AA; 38732 MW; B45C8C35B20ED24D CRC64;

Query Match 17.0%; Score 342; DB 1; Length 358;
Best Local Similarity 28.2%; Pred. No. 4.4e-19;
Matches 110; Conservative 66; Mismatches 162; Indels 52; Gaps 16;

OY 1 MGSLVYTGNDIRYSE-TYPEPEIKNPNDVKIKVSYCGIGTDLKEFTYSGGFVPEPKG 59
DB 1 MALALAFGARFGYKTKIDPQP-MGCPEDVYVIFIKAAALCGADMKRYNDSC----- 51
OY 60 TKDKISGYELPLCPGHESFGTVVSGVTSVKPGDRVAVEATSHCSDSRKYKDTVAODL 119
DB 52 -SDEFNSIR-----GHEFAGCINAGVEKVKDMKVGQRYVSDNSGHV-----C 92
OY 120 GLCMACQSSPNCASLSCGGLGAS--GGFAEY-VYGE-----DHVKKLPDSIPDDI 170
DB 93 GCPACQEQDPLCTCEKVNIGLDNNTWGGGFSKYCLVPEIILKIRHRAIMEIPDGDYED 152
OY 171 GALVERISYAMVAH-ERARFOPGOTALVLCGGPIGLATILALOGHAKGIV--CSEPAL 226
DB 153 AAVLDPICAKYKSIAGQSFLEPGQDVYVIGTGLGFSQMARIMAGAVNIIVVGLQEDVA 212
OY 227 IRRQPAKELGA-EVEDPSTCDANAVLAKAVPENEGFHAAPDCSGVPQFTTSIVATGPS 285
DB 213 VEEPPAKELGATAVVGVSTEDVAVRCQICGKDNGL--VIECSGANIALKQAIEMLRN 270
OY 286 GIAVVAWAGDHPICGMPMSLTYQEKYATGSMCYTKDQEVVYKALDELGISDKARKKI 345
DB 271 GGVAVRGM-GEPRLDPSINDITAMNKSIIIGHMAVYSTWRMIIRLLASAI--KVKPMI 326
OY 346 TGVVHLKDGVEKGFOLIEKENNVKIIVT 375
DB 327 THRIGLSQWRE-GFDAMVD--KTAIKVIMT 353

RESULT 8
XYL2_PICST
ID XYL2_PICST STANDARD; PRT; 363 AA.
AC P22144;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-xylose reductase (EC 1.1.1.9) (Xylitol dehydrogenase) (XDH).
GN XYL2.
OS Pichia stipitis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.

OX NCBI_TaxID=4924;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 5774;
RX MEDLINE=91168296; PubMed=2127555;
RA Koeller P., Amore R., Hollenberg C.P., Ciriacy M.,
RT Isolation and characterization of the Pichia stipitis xylitol
RT dehydrogenase gene, XYL2, and construction of a xylose-utilizing
RT Saccharomyces cerevisiae transformant.";
RL Curr. Genet. 18:493-500(1990).
RN [2]
RP SIMILARITY TO OTHER ZINC-ALCOHOL DEHYDROGENASES.
RX MEDLINE=93279386; PubMed=8504864;
RA Persson B., Hallborn J., Walfridsson M., Hahn-Haegerdal B.,
RA Keranen S., Penttilae M., Joernvall H.;
RT "Dual relationships of xylitol and alcohol dehydrogenases in families
RT of two protein types.";
RL FEBS Lett. 324:9-14(1993).
CC -1- CATALYTIC ACTIVITY: Xylitol + NAD(+) = D-xylose + NADH.
CC -1- COFACTOR: ZINC (potential).
CC -1- PATHWAY: D-xylose degradation.
CC -1- INDUCTION: BY XYLOSE. REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC
DR EMBL: X55392; CAA9066.1; -
DR EMBL: A16166; CAA01269.1; -
DR PIR: S13529; S13529.
DR HSSP: P07846; 1SDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR002025; NAD_binding.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Xylose metabolism; NAD; Zinc.
FT METAL 41
FT METAL 41
FT METAL 66
FT METAL 66
FT METAL 159
FT METAL 159
FT NP_BIND 183
FT NP_BIND 188
SQ SEQUENCE 363 AA; 38521 MW; 39E16D087160248 CRC64;

Query Match 17.0%; Score 342; DB 1; Length 363;
Best Local Similarity 28.4%; Pred. No. 4.5e-19;
Matches 110; Conservative 56; Mismatches 159; Indels 62; Gaps 14;

OY 10 NDIRSETVEPEIKNPNDVKIKVSYCGIGTDLKEFTYSGGFVPEPKGTGDKISGYEL 69
DB 13 DDISF-ETVDAPRISPTDVLVQVKKTGICGSDIHYFANG-----RIGNFVL 58
OY 70 -PLCGHESFGTVVSGVTSVKPGDRVAVEATSHCSDSRKYKDTVAODLGLCACQS 127
DB 59 TKPMVLGHESAGTVVQVGVTSKLVGDVNALEP--GIPRSFSEY-----KS 104
OY 128 GSPNCCASLSPFC-----GLGASGGAFAEYVYVGGDHVWKLPDSIPDDIGALVEPISA 180
DB 105 GHYNLCPHMAFAATPKSRKGEPRPTCLCKYKSPEDFLVKKPDHVSLEGLALVEPLSAG 164
OY 181 WHAVERARPOGOTALVILGGPIGLATILALOGHAKGIVCSEPALIRROPAKELGA--E 238
DB 165 VHASKLGSAVAFDYAVVAFGAGVGLAAVAARTFGAKGVIVDIDPNKRLKMAKDGAATH 224
OY 239 VEDPSTCDANAVIKAM---VPENEGFAAPDCSGVPQFTTSIVATGSGGIAVVAWAG 295
DB 225 TFNSKT-GGSEELIKAFGGVNP-----NVLECTGAEPCIKLGVDAIAGRGFPVQGN-A 277


```

Db      148 LACTHAIKAKTSINKGDLVVTGPPGIGL---LAAOVANSHGTVIITGSLNDQVRLKKA 204
Qy      233 KELGAEVFPSPSTDNDANVAKAMVPE-NEGFAH-AFDCSGVPQFTTSLVATPGSGIAV 289
Db      205 KEGVIDY----AIDTQEVDIKELVSELTLDGADVYLECGGAPVAAKQGIIDLRRKGGQYA 260
Qy      290 NVAVMWDHPHIGFPMPSLITQEKATQSMCTYVADPQEVYKALDGLSLDKAKMTTKGY 349
Db      261 QVGLFQAPQELQEFNEFKITQIEISVSGSRQPADMEPALSLNEKKRVN--AKTLVTHEY 317
Qy      350 HMKDGEKGFQKQIEHKENNVKILVLP 376
Db      318 TISEW-DKAYVHAH-KSGEAIKVLPLP 341

RESULT 11
YDYL_ECOLI
ID      YDYL_ECOLI      STANDARD;      PRT;      347 AA.
AC      P77280; P78174;
DT      01-NOV-1997 (rel. 35, Created)
DT      01-NOV-1997 (rel. 35, Last sequence update)
DT      16-OCT-2001 (rel. 40, Last annotation update)
DE      Hypothetical zinc-type alcohol dehydrogenase-like protein ydJf.
GN      ydJf or B1774.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxId=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12 / MG1655;
RX      MEDLINE=97426617; PubMed=9278503;
RA      Blattner F.R., Plunkett G. III, Bloch C.A., Berna N.T., Burland V.,
RA      Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12."
RL      Science 277:1453-1474(1997).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-K12.
RX      MEDLINE=97251357; PubMed=9097039;
RA      Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA      Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA      Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA      Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshida T., Saito N.,
RA      Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
RA      Takekoshi K., Takeuchi Y., Wada C., Yamamoto Y., Horinuchi T.;
RT      "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT      corresponding to the 28.0-40.1 min region on the linkage map.";
RL      DNA Res. 3:363-377(1996).
CC      -1- COPACITOR: ZINC (Potential).
CC      -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC      FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
DR      EMBL; AE000272; AAC74844.1; -
DR      EMBL; D90820; BAAL1565.1; -
DR      EMBL; D90821; BAAL1572.1; -
DR      HSSP; P07846; ISDG.
DR      EcoGene; EGI3486; ydJf.
DR      InterPro; IPR002328; ADH_zinc
DR      InterPro; IPR002085; Adh_zn_family.
DR      InterPro; IPR000205; NAD_binding.
DR      Pfam; PF00107; adh_zinc; 1.
DR      PROSITE; PS00059; ADH_ZINC; 1.

```

Hypothetical protein: Oxidoreductase; zinc: Complete proteome.
 FT METAL 39 39 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 65 65 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 98 98 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 101 101 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 109 109 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 152 152 ZINC (CATALYTIC) (BY SIMILARITY).
 SQ SEQUENCE 347 AA; 37701 MW; F8E2172184626A3B C8C64;

Query Match 16.7%; Score 337; DB 1; Length 347;
 Best Local Similarity 30.6%; Pred. No. 1e-18;
 Matches 113; Conservative 51; Mismatches 149; Indels 56; Gaps 14;

DB 8 GTNDRYSETPPEIKPNNDVIRKYSYCGICGTDLKEPTYSGVFPFKGKTDKISCY 67
 12 GTMKI-ISEIPIVKR---EDELIVKEVYIGSDVHGF-ESGFPPIPKDNPQ----- 59
 OY 68 ELPLCPGHEFGSTVEVSGVTSVKPGDRAVEATSHCSDRSRYKDYVAQDLGLMACQS 127
 DB 60 EIGL-GHECACTVAVGSRVAKFKPGDRVNIPEGVPC-----GHCRCYCLE 103
 OY 128 GSPNCCASISFGCL-GGASGGAEEVYVGEDHMKLPDISIPDDIGALVEPISVAMHAEV 186
 DB 104 GKYNICPDVDFMATQPNYKRGALTLYLCHPESTFKLPDMMDMEGALVEPAVGMHAAVL 163
 OY 167 ARFQGGTALVGGPGIGATLALQGHNAKIVCSEPALIRQPAKELGAEPDSTCD 246
 DB 164 ADVKPGKTIILGAGCIGLMTLQACKCLGATEIAVVDLEKLAEMEOLGATVINGAKE 223
 OY 247 DANAALKAMVPENEGFAHAFDCSGVPQETTS--IVATGPGSIAVAVAGDHPIGFMPM 304
 DB 224 DTIAACQGET-EDMGADIVFETAGSANTYKQAPLYMRG-GKIMIVGYTPGDSAINFLKI 281
 OY 305 S-----LTYQEKY-----ANG-----SMCYTKDQEVNALEDGLISLDRARM 344
 DB 282 NREVTIQTFVFRANRYPTVTEAISGREDPKSMVTHIYRDVQDAFEE---SVNKRDI 338
 OY 345 ITGKVLMD 353
 DB 339 IKGVIKISD 347

RESULT 12
 DHSO_SHEEP STANDARD; PRT: 354 AA.
 AC P07846;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sorbitol dehydrogenase (EC 1.1.1.14) (L-Iditol 2-dehydrogenase).
 GN SORD.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OX NCBI_TaxID:9940;
 RN [1]
 RN [1]
 RP TISSUE=Liver;
 RC MEDLINE=84158607; PubMed=6705798;
 RA Jeffery J., Cederlund E., Joernvall H.;
 RT "Sorbitol dehydrogenase. The primary structure of the sheep-liver
 RT enzyme.";
 RL Eur. J. Biochem. 140:7-16(1984).
 RN [2]
 RP SIMILARITY TO ZINC ALCOHOL DEHYDROGENASES.
 RX MEDLINE=84158592; PubMed=6368230;
 RA Joernvall H., von Bahr-Lindstrom H., Jeffery J.;
 RT "Extensive variations and basic features in the alcohol
 RT dehydrogenase-sorbitol dehydrogenase family.";
 RL Eur. J. Biochem. 140:17-23(1984).
 RN [3]

RP 3D-STRUCTURE MODELING.
 RA MEDLINE=86131624; PubMed=2936393;
 RX Eklund H., Horjales E., Joernvall H., Branden C.I., Jeffery J.;
 RT "Molecular aspects of functional differences between alcohol and
 RT sorbitol dehydrogenases.";
 RL Biochemistry 24:8005-8012(1985).
 CC -1 CATALYTIC ACTIVITY: L-Iditol + NAD(+) = L-sorbose + NADH.
 CC -1 COFACTOR: ZINC.
 CC -1 SUBUNIT: HOMOTETRAMER.
 CC -1 SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC PIR: S10065; S10065.
 DR PDB: 1SDG; 15-OCT-94.
 DR InterPro: IPR002328; ADH_zinc.
 DR InterPro: IPR002085; Adh_zn_family.
 DR InterPro: IPR000205; NAD_binding.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_zinc; 1.
 KW Oxidoreductase; zinc; NAD; Acetylation; 3D-structure.
 FT MOD_RES 1 1
 FT METAL 43 43 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 67 67 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 153 153 ZINC (CATALYTIC) (BY SIMILARITY).
 FT STRAND 4 11
 FT TURN 14 15
 FT STRAND 19 25
 FT TURN 30 31
 FT STRAND 32 41
 FT TURN 44 51
 FT HELIX 52 52
 FT TURN 57 58
 FT STRAND 63 64
 FT STRAND 69 76
 FT TURN 78 79
 FT TURN 85 86
 FT STRAND 88 91
 FT HELIX 101 103
 FT TURN 104 104
 FT TURN 112 113
 FT TURN 116 117
 FT STRAND 127 127
 FT STRAND 131 133
 FT HELIX 134 136
 FT STRAND 137 139
 FT TURN 146 148
 FT TURN 149 152
 FT HELIX 153 153
 FT TURN 154 163
 FT TURN 164 166
 FT TURN 169 170
 FT STRAND 172 176
 FT HELIX 180 191
 FT STRAND 196 200
 FT HELIX 204 212
 FT TURN 213 214
 FT STRAND 217 219
 FT HELIX 221 223
 FT TURN 228 235
 FT TURN 236 238
 FT STRAND 242 245
 FT HELIX 250 258
 FT TURN 259 259
 FT TURN 262 264
 FT STRAND 266 269
 FT TURN 282 283
 FT STRAND 287 289
 FT TURN 292 293
 FT HELIX 297 309
 FT TURN 310 311
 FT HELIX 316 318
 FT STRAND 319 321
 FT STRAND 324 324
 FT HELIX 325 335

01-OCT-1994 (Rel. 30, Created)
01-OCT-1994 (Rel. 30, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Starvation sensing protein rspB (EC 1.1.1.-).
RSPB OR B1580.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
[1]
SEQUENCE FROM N.A.
STRAIN=K12 / W3110;
MEDLINE=94310441; PubMed=7545940;
Huisman G.W., Koller R.;
"Sensing starvation: a homoserine lactone-dependent signaling pathway in Escherichia coli.";
Science 265:537-539(1994).
[2]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
"The complete genome sequence of Escherichia coli K-12.";
Science 277:1453-1474(1997).
[3]
SEQUENCE FROM N.A.
STRAIN=K12;
MEDLINE=97251357; PubMed=9097039;
Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., Kasai H., Koshimizu K., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N., Sakemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
"A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map.";
DNA Res. 3:363-377(1996).
-1- FUNCTION: NOT KNOWN; PROBABLE CATABOLIC ENZYME.
-1- CORFACTOR: ZINC (Potential).
-1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L31628; AAA21686.1; -
EMBL: AE000254; AAC74652.1; -
EMBL: D90799; BAA15284.1; -
EMBL: D90800; BAA15306.1; -
Ecogene: EG12452; rspb.
InterPro: IPR002328; ADH_zinc.
InterPro: IPR002085; Adh_zn_family.
Pfam: PF00107; adh_zinc; 1.
PROSITE: PS00059; ADH_ZINC; 1.
Oxidoreductase; zinc; Complete proteome.
METAL 37 37 ZINC (CATALYTIC) (BY SIMILARITY).
METAL 59 59 ZINC (CATALYTIC) (BY SIMILARITY).
METAL 89 89 ZINC (SECOND ATOM) (BY SIMILARITY).
METAL 92 92 ZINC (SECOND ATOM) (BY SIMILARITY).
METAL 95 95 ZINC (SECOND ATOM) (BY SIMILARITY).
METAL 103 103 ZINC (SECOND ATOM) (BY SIMILARITY).
METAL 144 144 ZINC (CATALYTIC) (BY SIMILARITY).
SEQUENCE 339 AA; 36564 MW; E0A1062DB7C14E CMC64;

Best Local Similarity 27.2%; Pred. No. 2.2e-17;
Matches 107; Conservative 57; Mismatches 159; Indels 71; Gaps 16;

```

QY      1 MGGLYYGINDIRYSETVPEPEIKPN--DVKIKVSYCGICGTDLKEFTYSGGVPFPKQ 58
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MKSILIEKPNQL-----AIVEREIPPSAGEVRYKVLGICGSD--SHYRGHNPF----- 50
QY      59 GTMDKISGVELPLCPGHEPFGTVVEVSGVTSYKPGDRVAVAEATSHCSDSRYKCDTYAOD 118
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      51 -----AKYPRVIGHEFEFGVIDAVGEGVESARVGERVAVDPVSC----- 89
QY      119 LGLCMACQSGSPNCASLSFCGLGASGGFAEYVYGEDHMKLPDPSIPDDIGALVEPIS 178
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      90 -GHCPYCSIGKPNVCTTIAVLGV-HADGFSYAVYPAKNAKMIPEAVADQYAVMIEPFT 147
QY      179 VANHAVERARFQPGQTAIVLGSGPILATLALQG-HHAGKIVCSEPALIRQFAKELGA 237
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      148 IAAVNTGHGQPTENDIVLYYGAGPIGLITIVQLKGVYNNKNIIVADRIDERLEKAKESGA 207
QY      238 EVFDPSTCDANAVLKAMVPENEG-----HAFFDCSGVPQTFTTSIYATGSPGIAYN 290
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      208 DW-----AINNSQIPLGIEFTTEKGIKPTLIIDAACHPSILKEAVTLASP---AAR 254
QY      291 VAVWG--DHPIGFMPMSLTYQEKYATGSMCYTKDFQEVVKALEDGLISLDKARKMITGK 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      255 IVLMGFSSESEVYIQGIGITGKELISFSSRLNANK-PRIVIDWLSKGLI---KPEKLITHT 310
QY      349 V---HKDGYEKGFQKQLEHKENN-VKILVTPNE 378
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      311 FDFQHVADAI-----SLFEODQKHCKCKVLLTFSE 339

```

Search completed: May 2, 2003, 12:29:29
Job time : 287 secs

GenCore version 5.1.4-P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 10:19:42 ; Search time 1285 Seconds
(without alignments)
60.932 Million cell updates/sec

Title: us-10-020-674-2

Perfect score: 2013
Sequence: 1 MKGLLYGTNDIRSETVPE.....QLIEHKENVKILVTPEVS 380

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_proteint:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	588.5	29.2	349	2	O93R65
2	571	28.4	348	16	O9K5Y6
3	567.5	28.2	363	16	O9HM8
4	566.5	28.1	362	2	O59696
5	551.5	27.4	346	16	O34788
6	538.5	26.8	370	16	O8CH42
7	537	26.7	354	16	O9K0J3
8	534	26.5	354	16	O9JYJ8
9	455.5	22.6	352	16	O9HWT1
10	422.5	21.0	397	16	O8U6R7
11	414.5	20.6	352	5	O96496
12	386.5	19.2	347	5	O21702
13	386.5	19.2	353	3	O74230
14	370	18.4	395	16	O9WYR3
15	369	18.3	339	16	O35045
16	368.5	18.3	347	5	O21703

17	365.5	18.2	357	3	O07786	O07786 saccharomyc
18	365	18.1	371	10	O9MAN7	O9MAN7 eritobrya
19	364.5	18.1	348	16	O92EF4	O92EF4 listeria in
20	364	18.1	371	10	O8W2D0	O8W2D0 malus domes
21	361	17.9	347	16	O92P23	O92P23 rhizobium m
22	361	17.9	371	10	O92R22	O92R22 malus domes
23	358.5	17.8	348	16	O8Y9M0	O8Y9M0 listeria mo
24	355.5	17.7	360	5	O96299	O96299 drosophila
25	352.5	17.5	354	16	O9KGB7	O9KGB7 bacillus ha
26	352.5	17.5	360	5	O97479	O97479 drosophila
27	352	17.5	367	10	O9MBD7	O9MBD7 prunus pers
28	350	17.4	368	10	O93X81	O93X81 prunus cera
29	349.5	17.4	360	5	O960H1	O960H1 drosophila
30	348	17.3	343	16	O927H6	O927H6 listeria in
31	348	17.3	343	16	O8Y414	O8Y414 listeria mo
32	346	17.2	338	16	O8ZM59	O8ZM59 salmonella
33	346	17.2	371	16	O8UFI9	O8UFI9 agrobacteri
34	345	17.1	358	16	O8XD08	O8XD08 escherichia
35	343	17.0	343	16	O930C9	O930C9 rhizobium m
36	342	17.0	289	16	O98CF7	O98CF7 rhizobium l
37	340	16.9	344	16	O8XR80	O8XR80 raietonia s
38	339	16.8	343	16	O8U8K4	O8U8K4 agrobacteri
39	338	16.8	371	16	O8ZBQ5	O8ZBQ5 yersinia pe
40	337.5	16.8	389	16	O8YFE3	O8YFE3 anabaena sp
41	337	16.7	368	10	O8W2C9	O8W2C9 malus domes
42	336	16.7	368	10	O8W2C8	O8W2C8 malus domes
43	333.5	16.6	356	3	O07993	O07993 saccharomyc
44	333.5	16.6	357	6	O18769	O18769 callithrix
45	331.5	16.5	348	17	O9UYX0	O9UYX0 pyrococcus

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	349 AA.
ID O93R65			
AC O93R65:			
DT 01-DEC-2001 (TREMBLrel. 19, Created)			
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Acetylacetoalcohol reductase.			
GN ACR.			
OS Bacillus cereus.			
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC Bacillaceae; Bacillus.			
OX NCBI_TaxID=1396;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Hosaka T., U.S., Ohkuma M., Kudo T.;			
RT "Characterization of the NADH-linked Acetylacetoalcohol Reductase / 2,3-			
RL J. Biosci. Bioeng. 0:0-0(2001).			
DR EMBL; AB063194; BAB60856.1; .			
DR InterPro: IPR002328; ADH_zinc.			
DR InterPro: IPR002085; Adh_zn_family.			
DR Pfam: PF00107; adh_zinc.1.			
DR PROSITE: PS00059; ADH_ZINC; UNKNOWN.1.			
SQ SEQUENCE 349 AA; 37720 MW; 87FAA979176113E1 CRC64;			
Query Match	29.2%;	Score 588.5;	DB 2; Length 349;
Best Local Similarity	39.4%;	Pred. No. 1.6e-37;	
Matches 147;	Conservative 58;	Mismatches 139;	Indels 29; Gaps 12;
OY 1 MKGLLYGTNDIRSETVPEPEIKNDYKIVSYGICGTDLKEFTYSGGVPPFKQST 60			
DB 1 KMALMHNDRDYVEE-VPEPTVK-PGAVKIKYKMGICGTDLHE--YLAGFIFITE-- 54			
OY 61 KKKISYEPLCPGHEFSGTVVEVSGSVYPCGRVAVEAHSCHSDRSRYKDYVQDQD 120			
DB 55 EHPDTHVAKPVLIGHEFSEVEVEIGVTSKVGDMVVEPIYSC-----G 100			
OY 121 LCMACGSGSPNCASLSFCGLGASGFAEYVVYGGEDHMKLPDSIPDIALVEPISYA 180			

Db 101 KCACGCHGHVNWCEQVLFHIGSGEGGFSFYTVPPEDMWHHHPDEKTYTDOGALVEPAAVA 160
Qy 181 WHAVEARPOGOTATLVGGPGLATITLALOGHNAKIVCSEPALIRROFAKELGAE-V 239
Db 161 VHAVROSKLEGEAVAFECGPGILGLVIOAKAAGATPIYAVELSELREBLAKLAGADYV 220
Qy 240 FPDSTDDANAVLKKANVPENEGFHAFCDSGVPOTFTTSIVATGSGIANNVAVMGDHP 299
Db 221 LNPAT-QDVLAEIRLNT-NGLVNWSFEVTGVEVLRQAIESTSEFGTIVISVW-EKDA 277
Qy 300 GEMPSLTYOEKATGSMCYTAKDFOEYKALEDLISDKARKMTGKVHLKDGKECF 359
Db 278 TTPNNLVYKEKEVGILGFR-HIPPAVKILISSQI---QAEKLIKTIYDQVEEGE 333
Qy 360 KOLIEKKNVAKI 372
Db 334 EALVVKD-TQVKI 345

RESULT 2

ID O9K5Y6 PRELIMINARY; PRY; 348 AA.
AC O9K5Y6;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE L-Iditol 2-dehydrogenase.
GN BH3949.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RA MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC - COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY
DR EMBL: AB001520; BAB07668.1; -
DR HSSP: P07846; ISDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; zinc; Complete proteome.
SQ SEQUENCE 348 AA; 37516 MW; 28146770BE77FEED CRC64;

Query Match 28.4%; Score 571; DB 16; Length 348;
Best Local Similarity 37.7%; Pred. No. 3 6e-36;
Matches 141; Conservative 65; Mismatches 134; Indels 34; Gaps 13;

Qy 1 MKGLIYGTNDIYSETVEPEIKNPNDYKIVSYCGICGTDLKETTYSGGVPFFPKQGT 60
Db 1 MQLRMHGVADVAVDDIAEPATLK--GKYKIKYKMGICGSDLHE--YAGPIFIF-QNS 55
Qy 61 KDIISGYELPLCGHEFSGTVEVSGSVTSVRCGDRVAVNATSHCSDRSRKYKTVAQDLG 120
Db 56 PHLTLEKAPIYVGHFSGGVVGVGATYKCEGDRVAVVEPIFAC-----G 101
Qy 121 LCMACGSGPNCASISFCGLGASGFAEYVYVYGEDHNV-KLPDSIPDDIGALVEPISV 179
Db 102 TCYACGKGYKNCCEQGLGIGAGGGGSEYIVY-DEHNVKIPDVVSFGGLVPEPAV 160
Qy 180 AHAVARARPOGOTATLVGGPGLATITLALOGHNAKIVCSEPALIRROFAKELGAEV 239

Db 161 ALYAVROSQFNVDQAVVGTGPGLITLALKASGATYAVELSTERRORAEQDLGATA 220
Qy 240 FPDSTDDANAVLKKANVPENEGFHAFCDSGVPOTFTTSIVATGSGIANNVAVMGDHP 299
Db 221 LNPAT--EVNVAEEIORTLDGADVSEFTGVPIYLTQAIESTKINGOTMVISI-EKEA 276
Qy 300 GEMPSLTYOEKATGSMCYTAKD-FOEYKALEDLISDKARKMTGKVHLKDGKECF 358
Db 277 SFQPNLVYKRENNIGITCY--KDLFPAVISLMNGYFS--AETLVTKRIGLHDIVBOG 331
Qy 359 EKOLIEKEN-NVK 371
Db 332 FEALM-KTNCNIK 343

RESULT 3

ID O9HWM8 PRELIMINARY; PRY; 363 AA.
AC O9HWM8;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE 2,3-butanediol dehydrogenase.
GN PA4153.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RA MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Garbet R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen";
RL Nature 406:959-964(2000).
CC - COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY
DR EMBL: AE004831; AAG07540.1; -
DR HSSP: P07846; ISDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR00205; NAD_binding.
DR Pfam: PF00107; adh_zinc; 1.
DR PROSITE: PS00059; ADH_ZINC; 1.
KW Oxidoreductase; zinc; Complete proteome.
SQ SEQUENCE 363 AA; 38193 MW; 3B4DD83ACA1027B8 CRC64;

Query Match 28.2%; Score 567.5; DB 16; Length 363;
Best Local Similarity 36.1%; Pred. No. 7 2e-36;
Matches 136; Conservative 73; Mismatches 141; Indels 27; Gaps 11;

Qy 1 MKGLIYGTNDIYSETVEPEIKNPNDYKIVSYCGICGTDLKETTYSGGVPFFPKQGT 60
Db 11 MRAAVHGHGHDIV-EDVLPAPPPGWQIRVHMGICGSDLHE--YLAGVPIFVE-A 66
Qy 61 KDIISGYELPLCGHEFSGTVEVSGSVTSVRCGDRVAVNATSHCSDRSRKYKTVAQDLG 120
Db 67 PHLTGLKQDCLTGHFSESEIYRLNGVYGFVAGVAVADAQCHC-----G 112
Qy 121 LCMACGSGPNCASISFCGLGASGFAEYVYVYGEDHNV-KLPDSIPDDIGALVEPISV 180
Db 113 TCYCRHGHLYNCENIAFTGLMN-NGAFAYVNVANPLLYALPAGFPSPGALILEPLAVG 171
Qy 181 WHAVEARARPOGOTATLVGGPGLATITLALOGHNAKIVCSEPALIRROFAKELGAEV 239
Db 172 MHAVKAGSLGQNVVYVAGATIGLSTIMCARAAGAAOVIALEMSSARKAKALEVGAQOV 231

QY 240 FDPSCDANAVIKAMVPENEGEHAAPDCSGVPOTFTTSIVATGPGSIANVAVMGDHP1 299
 DB 232 LDPSC-DALGELIRALT-GGLGADVSFECTGNKHAKLIDIRKAGKCVLVGIF-EEPS 288
 QY 300 GFMPSLTYOEKATGSMCTYTKDFOEYVAKLEDGLISLDKARKMTTGKVLKDGVEKGF 359
 DB 289 EFNFEVLSTERQQLGALAYN-GEFADYIAFTADGRLOI---APLVTRIGLEIYERGF 344
 QY 360 KOLIEHKENNVKILVTP 376
 DB 345 EELVNNKEHNKIVSP 361

RESULT 4

QY 059696 PRELIMINARY: PRT: 362 AA.
 AC 059696:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE 2,3-butanediol dehydrogenase (EC 1.1.1.4).
 GN ADH.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 CX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PPG2;
 RX MEDLINE=95113288; PubMed=7813883;
 RA Huang M., Oppermann F.B., Steinhuebel A.;
 RT "Molecular characterization of the Pseudomonas putida 2,3-butanediol
 RT catabolic pathway." Lett. 124:141-150(1994).
 RL FEMS Microbiol. Lett. 124:141-150(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PPG2;
 RA Oppermann F.B.;
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 DR EMBL: L35343; AAS58982.1; -
 DR HSSP: P07846; 1SDG.
 DR InterPro: IPR002328; ADH_zinc.
 DR Pfam: PF00107; adh_zinc.1.
 DR PROSITE: PS00059; ADH_ZINC.1.
 KW Oxidoreductase; Zinc.
 SQ SEQUENCE 362 AA: 38386 MW: 503CD0A58C8BB816 CRC64:

Query Match 28.1%; Score 566.5; DB 2; Length 362;
 Best Local Similarity 36.6%; Pred. No. 8.5e-36;
 Matches 138; Conservative 65; Mismatches 147; Indels 27; Gaps 11;

QY 1 MGGLLYGNDIRKSTVEPEPKNNNDYKIKSYCGICGTDLKETYSGGPFPFKOGT 60
 DB 9 MRAAVHGNNDIRV-EQVPLPADPARGWQIKYDMCGICGSDIHE--YVAGPFIIVE-A 64
 QY 61 KOKISGEYELPLCGHEFSTVVEGSGVTSVKGRDRAVEATSHCSRKYKQTVADG 120
 DB 65 PRLPTIGOCQILGHEFCQIAKLGSGVEGFANVDVADACHC-----G 110
 QY 121 LCMACOSSPNCASISFCGLGASGFAYEYVYGGEDHNVKLPDIPDDIGALVERISA 180
 DB 111 TCYUCTHGLYICERLAFGLIMN-NGAFALVNVNPNLLYRLPGFPPEAGALIEPLAVG 169
 QY 181 WHAVEARARPOGTALVLGGPRTGLATIALOGHNAGKIVCSRPALIRQFAELGAE-Y 239
 DB 170 MIAVKAAGSLGQTVVAVAGTIGTICTICAKAAGAAVATLEMSKARAKAVEATV 229
 QY 240 FDPSCDANAVIKAMVPENEGEHAAPDCSGVPOTFTTSIVATGPGSIANVAVMGDHP1 299

DB 230 LDPSC-DALGELIRALT-FGLGADVSFECTGNKHAKLIDIRKAGKCVLVGIF-EEPS 286
 QY 300 GFMPSLTYOEKATGSMCTYTKDFOEYVAKLEDGLISLDKARKMTTGKVLKDGVEKGF 359
 DB 287 EFNFEVLSTERQQLGALAYN-GEFADYIAFTADGRLOI---RPLVTGRIGLEQIVEIGF 342
 QY 360 KOLIEHKENNVKILVTP 376
 DB 343 EELVNNKEHNKIVSP 359

RESULT 5

QY 034788 PRELIMINARY: PRT: 346 AA.
 AC 034788:
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE DEHYDROGENASE.
 GN YDLI.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 CC Bacillaceae; Bacillus.
 CX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WARBURG 168;
 RA Kasahara Y., Nakai S., Ogasawara N., Yata K., Sadate Y.;
 RL DNA Res. 0:0-0(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertiero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Eutlan K.D., Errington J., Fabre C., Ferrati E., Fougere D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kjaer-Blandhard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tostato V., Uchiyama S., Vandendol M., Vanlier F., Vassarotti A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Welzenoeger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yatsunaga K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis." Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.

```

DR   EMBL: AB007638; BAA22767.1; -
DR   EMBL: J299107; CAB12443.1; -
DR   HSSP: P07846; 1SDG.
DR   InterPro: IPR002328; ADH_zinc.
DR   InterPro: IPR002085; Adh_zn_family.
DR   Pfam: PF00107; adh_zinc.1.
DR   PROSITE: PS00059; ADH_ZINC_1.
KW   Oxidoreductase; Zinc; Complete proteome.
SQ
SEQUENCE 346 AA; 37341 MW; 5E327192678F8A2 CRC64;

Query Match 27.4%; Score 551.5; DB 16; Length 346;
Best Local Similarity 38.4%; Pred. No. 1,2e-34;
Matches 147; Conservative 55; Mismatches 138; Indels 43; Gaps 13;

QY 1 MGLIYGTGNDIRSEYEPPEIRKPNNDYKIKVSTGICIGTDLKKEFTTSGGFVFPKQGT 60
DB 1 MKAARHNMOKDIRI-EHIEEPKTE-PGKVKIKVKKCGICGSDLHE--YLGDFIFP--- 52
QY 61 KDK---ISGYELPLCPGHFSGTVEVSGVTSVPGGRVANA--TSHCSDRSRYKDTV 115
DB 53 VDKRPLNERNPVPYMGHEFSEVYEVEEGVENVYGRDVRVVEPFLFANH----- 101
QY 116 AODLGLCAACOSGSPNCASUSFCGLGASGGFAEYVYVYGEDHMYKLDSIPDDIGALVE 175
DB 102 -----GHGGAYNLDEOMGFLTLAAGGGGSEYVSVDDELFLKLPDELSEOGALVE 152
QY 176 PISVMHVERARPPGGOTALVYGGGPGIGLMTIALOGHNAKGIYCSFPALIRFPAEL 235
DB 153 PSAVALIVRRSKLAKGKAAPFGCGPIGLVLEALKAGAADIVAYELSPERDOKABEL 212
QY 236 GAEPDPESTCDANAVLKAAMPENEGFAAFDCSGVPOTFTTSTIATVPSGIANVAWVG 295
DB 213 GAIYDPEKTDVVAEIAERT--GGGVDAVEVNGVPVYLQAIQSTIIAGEYIVLSW- 269
QY 296 DHPIDFMMSLTQOKKATGSMCTYVKP-FOEYKALDELGLSLDKARKMTIGVYHLKDG 354
DB 270 EKGAEIHFNDVIYKRIYKGIIGY--RDIFFAVLSLMKEGYFSAD--LKVTKKIVLDL 324
QY 355 VEKGFQOLIEHKENNVKILVTPN 377
DB 325 IEEGFGALIKER-SQVKILTVRPN 346

RESULT 6
Q9CH42 PRELIMINARY; PRT; 370 AA.
ID AC Q9CH42;
DR 01-JUN-2001 (TREMBLrel. 17, created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE 2,3-butanediol dehydrogenase (EC 1.1.1.4).
GN BTDB OR L10897.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OX Streptococcaceae; Lactococcus.
NCBI_TaxID=1360;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mungier S., Jajillon O., Malarne K.,
RA Weissensbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL GEMOME Res. 11:731-753(2001).
CC -1 COFACTOR: RQDIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -1 SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL: AE006323; AAK04995.1; -
DR HSSP: P07846; 1SDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc.1.
DR
```

DR PROSITE, P500059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc; Complete proteome.
SQ SEQUENCE 370 AA; 39445 MW; D53348B51D8030DB CRC64;

Query Match 26.8%; Score 538.5; DB 16; Length 370;
Best Local Similarity 37.3%; Pred. No. 1,3e-33;
Matches 144; Conservative 53; Mismatches 144; Indels 45; Gaps 14;

QY 1 MGLLVYGTNDIRYSETVEPEIKPNNDYKIKVSYCGIGTDLKEFTYSGGFVEPPKGT 60
DB 15 MRARFYDRGDRIIDE-INEPIYK-AGQGVIDVAMCGICGTDLHEFL--DGFIFCPSAEH 70
QY 61 KKKISYEELPLCPGHEFSGSTVVEGSGYTVKPGGRVAVE--ATGHCDSRKYKNTVAOD 118
DB 71 PNPITGVPPVTLGHMSSVNVNIGEGVSGLKGSHVVEPYEYIPEGDITD----- 121
QY 119 LGICMAQSGSNNCASISFCGLGASGGFAEYVYGGEDHMKLDPDISPDIGALVEIS 178
DB 122 -----ELGHYNLBSGNSFGLGGNGGLAKISVDENWYHKIPDNLPLDEALILEPLS 174
QY 179 VAMHAYERAFRPGQTALVVGSGPIGL-ATTLALQGHAGKIYCSSEPALIRPOFAKLG 236
DB 175 VGYHAYERANLSEKSTVLYVAGPIGLTLAAVAKAOGH---TVIISPSGLRRKKAQE-- 229
QY 237 AEV----PDPSCTDDANAVLKAMVE--NEGFAVFDGSGVDPQPTTSIVATGSPGIAYN 290
DB 230 AOVADYFFNPIDD-----IQAKVHEINEKGYDAFAECTSVOPGDACLDAILRMGGTYVI 284
QY 291 VAVMGDHPHPIGFMPSILTYOEKYATGSMCTYKDFQEVYKALEDDILSDKARKMITGKVH 350
DB 285 VAIWG-KPASYVMALKVIEKANLICTIAYN-NTHPKTIIDVSTGRIKID--QFTTAKG 339
QY 351 LKDVGEKGEKOLEHKENNVKLTMP 376
DB 340 LDDLIDKGFDTLIHNETAVKLTISP 365

RESULT 7
Q9K0J3 PRELIMINARY: PRT: 354 AA.
AC Q9K0J3:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DEF Alcohol dehydrogenase, zinc-containing.
GN MMB0604.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX MCB1_taxid=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.C.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blat E., Clifton H., Clark E.B.,
RA Causton M.D., Uitterback T.R., Khouri H., Qin H., Yamanehan J.,
RA Gill J., Scarlato V., Masigiani V., Piazza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58".
RL Science 287:1809-1815(2000).
CC -! COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -! SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL: AE002416; AAF1031.1; -
DR HSSP: P07846; 1SDG.
DR TIGR: MMB0604; -
DR InterPro: IPR002328; ADH_zinc.
DR Interpro: IPR002085; adh_zn_family.
DR Pfam: PF00107; adh_zinc; 1

SEQUENCE 352 AA; 38164 MW; 56360D3D2EDA6D26 CRC64;
Query Match 20.6%; Score 414.5; DB 5; Length 352;
Best Local Similarity 29.0%; Pred. No. 4.8e-24;
Matches 113; Conservative 65; Mismatches 140; Indels 71; Gaps 10;
7 YGTNDIRYSE-TYPEPEIKNPNDVIKIVSYCGICGTDLKEFTYSGGVFFPKGTCKDIS 65
11 YKQNDLLEQRPIPEPK---EDEVLLQMAVYIGSGSDVHYHGRIDFIYKD----- 60
66 GYELPLCPGHEFSGTVVEVSGVTSVKRQDRVAVENTSHCSDRSRYKDTVAODLGLCMAC 125
61 ---PMVIGHESGTVVYKKNVHLKKGDRVAVEPGVPCR-----RCQFC 102
126 QGSPNCCASLSEFCGLGASGFAEYVYGGEDHMKLPDIPDDIGALVEPISVAMHAVER 185
103 KEGRYNLCPPDLTFCATPPDDGNLARYVHADPCHKLPDNNVSLLEGALLPESVGHACR 162
186 RARPGQTALVLCGGPIGLATILALOGHNAGIKVSEPALIRROFAKELGAE---VDP 242
163 RAGVQLGTIVLIGAGPIGLVSLAAKAYGA-FVVCTRSPRRLVAVKNCADVTLYVDP 221
243 STCODANAVLKAMPENEGFAAFDCSGVPQFTTSIVATPSGIAVNAVWGDHPICFM 302
222 AKESSESTIERIRSAIGDLPVVTIDCSGENKICITIGINITRTGTLMLVGM----- 272
303 PMSLTQEKVATGSMCTYVK-----DFOEVVKALEDGLISD-----KARKMIT 346
273 -----GSOMVTVPLVNMACAREIDIKSVFRYCNPDYLALEMVASGRCNVQKOLVT 320
347 GKVLKDGVEKFKOLIEHKENNVKILVT 375
321 HSFLEQTV-D-AFEARKKADNTIKVMIS 348
RESULT 12
ID 021702 PRELIMINARY; PRT; 347 AA.
AC 021702;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE R0485.5 protein.
GN R0485.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology";
RL Science 282:2012-2018(1998).
CC -1-COFACITOR: ZINC (BY SIMILARITY).
DR HSSP; P07846; 1SDG.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; ADH_zn_family.
DR InterPro; IPR00205; NAD_binding.
DR Pfam; PF00107; adh_zinc1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc.
SEQUENCE 347 AA; 37323 MW; AE71536D4B1344B0 CRC64;
Query Match 19.2%; Score 386.5; DB 5; Length 347;
Best Local Similarity 29.6%; Pred. No. 6.9e-22;
Matches 112; Conservative 62; Mismatches 151; Indels 53; Gaps 12;

7 YGTNDIRYSETPPEIKNPNDVIKIVSYCGICGTDLKEFTYSGGVFFPKGTCKDISG 66
11 YGVDDLRN-EQVPIPK-PGRNQLVKNVHTVIGSGSDVHYHGAIGPVVKE----- 60
67 YELPLCPGHEFSGTVVEVSGVTSVKRQDRVAVENTSHCSDRSRYKDTVAODLGLCMAC 126
61 ---PMVIGHESGTVVYKKNVHLKKGDRVAMEGJLPCR-----LCENCK 103
127 QGSPNCCASLSEFCGLGASGFAEYVYGGEDHMKLPDIPDDIGALVEPISVAMHAVER 186
104 TGRYNLCPEMFEPATPPYHGLSRVHADPCHKLPDNNVSLLEGALLPESVGHACR 163
187 RARPGQTALVLCGGPIGLATILALOGHNAGIKVSEPALIRROFAKELGAEVDPSTCD 246
164 GNVGMHRYVLVIGAGPIGLVSLITAKVAGAKRVITDLDRLAKKLGAD---ATIN 219
247 DANAVLKAMPE-----NEGFHAFDCSGVPQFTTSIVATPSGIAVNAVWGDH--- 297
220 VKGKSLDAVKSEIITALDDQPDVCICTGAPSIETAITTTKSGGVIVLGLADRAVEI 279
298 PIGFPMSLTQEKVATGSMCTYVKDFOEVVKALEDGLISDARKMITGKVLHK-DGVE 356
280 PI---IESAGREVDMRGIERI-VNCPPTALELISGKLNLSGLT-----RAHFKLEETO 329
357 KGFOLIEHKENNVKILV 374
330 EAFKR--TQKADVIKVF1 345
RESULT 13
ID 074230 PRELIMINARY; PRT; 353 AA.
AC 074230;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Xylitol dehydrogenase (EC 1.1.1.9).
GN XDH.
OS Candida sp. HA167.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=78167;
RN [1]
RP SEQUENCE FROM N.A.
RA Habenicht A., Motejedded H., Wegerer A., Kies M., Mattes R.;
RT "Xylose utilisation. Cloning and characterisation of the xylitol
dehydrogenase from Galactocandida mastotermulis".
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1-COFACITOR: ZINC (BY SIMILARITY).
DR EMBL; AF072541; AAC24597.1.;
DR HSSP; P07846; 1SDG.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR00205; NAD_binding.
DR Pfam; PF00107; adh_zinc1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc.
SEQUENCE 353 AA; 37382 MW; 9BA7F10CE9F6687 CRC64;
Query Match 19.2%; Score 386.5; DB 3; Length 353;
Best Local Similarity 30.6%; Pred. No. 7.1e-22;
Matches 114; Conservative 48; Mismatches 161; Indels 49; Gaps 9;
11 DIRYSETPPEIKNPNDVIKIVSYCGICGTDLKEFTYSGGVFFPKGTCKDISG 70
16 DVKE-EDRPIKLPDPSYKIVKKTGTCGSDVHYHGHAIQDFYVK-----AP 63
71 LCPGHEFSGTVVEVSGVTSVKRQDRVAVENTSHCSDRSRYKDTVAODLGLCMACSGSP 130
64 MYLGHSSGVVLEVGSEVSKLVGDRVAMEPGVSRHSDYK-----SGHY 109
131 NCCASLSEFCGLGASGFAEYVYGGEDHMKLPDIPDDIGALVEPISVAMHAVERARPO 190

```

Db 110 NLCHEMFAATPPYDGLCKRYIIPEDEFCVLPHEVSLIEGALVEPLSAVHSSKLNK 169
QY 191 PGCFALVIGGGPGLATLIALQGHAGKIVCSEPALIRROFAKELGAEV-----FDSTC 245
Db 170 PGRVATVYAGCPVGLLVAAVASAGASVITIIDLVESRLNAKELGATATVQVDFKOTPK 229
QY 246 DDANAVILKAWPENEGF--HAAFDCSGVPOTFTTSIATGPGSIAVNAVAVMDHPIGFMP 303
Db 230 ESAKVVAA-----NNGIAPDVVIDASGAEASINSAINAIRPGTYVQVM-GKRDVSFPI 284
QY 304 MSLLYQKATNGSMCYVVKPQEVYKALDEGLISLDKARKMTTCKHLKQGV-----356
Db 285 ATLLGKELTVKGSFRYGYGDIPLAVSLASGVAV---KKLITHEVFEEDAAEAFLQVLD 341
QY 357 -KGFOLIEHKE 367
Db 342 GKAKCTIINGPE 353

RESULT 14
Q9WVP3
ID 09WVP3 PRELIMINARY; PRT; 395 AA.
AC 09WVP3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Alcohol dehydrogenase, zinc-containing.
GN TM0412.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_Taxid=2336;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL: AE001720; AAD35497.1; -
DR TIGR: TM0412; -
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000205; NAD_binding.
DR Pfam: PF00107; adh_zinc.1.
DR PROSITE: PS00059; ADH_ZINC.1.
KW Oxidoreductase; zinc; Complete proteome.
SQ SEQUENCE 395 AA; 43317 MW; 29D2A19F00946397 CRC64;

Query Match 18.4%; Score 370; DB 16; Length 395;
Best Local Similarity 34.0%; Pred. No. 1.6e-20;
Matches 109; Conservative 37; Mismatches 111; Indels 64; Gaps 13;

QY 4 LLYYGNDRYSEET--VPEEIKNPNDVKIKVSYCGICGNDL-KEFYSGGVPVPEPKQ 58
Db 27 LHWLGSKWRYREYREYEPPEPRIEKFTITIKVACGICGSDVHMAQTDDEEGYIYLP-- 84
QY 59 GKRDKISGYELPLDGHGHEFSGTVEVSGVTSVKP-----DRAVAEATSHCSDRSRK 112
Db 85 -----GLTGF--PYTLGHESSGVVVEAGPEAINRRTNKRREIEPVCHEMLMC----- 131
QY 113 DIVAADLGLCMACGSGSPCCASLSFCIGAGSGGFAEYV-----VYGED 157
Db 132 -----GHCRCACGAGFPNHCHENLNELGF-NVDGAFAEYVVDKAYAMSLRELEGYEED 183

```

```

QY 158 HMVKLPIDPIDGALVEPISVAMHA--VERARPGOTALVIGGGPGLATLIALQGH 215
Db 184 RLFL-----ACSLVEPTSVATNAVYIRGGIRPGDNVYLLGGPGLAAVILHAG 235
QY 216 AGKIVCSEPALIRROFAKELGAE-VFDPSTCDANAVILKAWPENEGFHAAFDCSGVPQT 274
Db 236 ASKYLSEPSFVRNRNAKELGADHDVIDPTKENFEAVLD--YTNGLGAKLFLEATVPQL 293
QY 275 FTTSI--VAIGPSCIANNVAV 293
Db 294 VMPQIEEVIWRARGINATVAI 314

RESULT 15
O35045
ID 035045 PRELIMINARY; PRT; 339 AA.
AC 035045;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Zinc-containing alcohol dehydrogenase.
GN YMD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Beterro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funas S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Ilaya M., Jones L.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate I.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serior P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi B., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosto V., Uchiyama C., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC -1- COFACTOR: REQUIRES ZINC FOR ITS ACTIVITY (BY SIMILARITY).

```

-1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR EMBL: AF015825; AAC46329.1; -
DR EMBL: 299110; CAB13090.1; -
DR HSSP: P07846; 1SDG.
DR InterPro: IPR002328; ADH_zinc.
DR InterPro: IPR002085; Adh_zn_family.
DR Pfam: PF00107; adh_zinc.1.
DR ProSite: PS00059; ADH_ZINC.1.
KW Oxidoreductase; zinc; Complete proteome.
SEQUENCE 339 AA; 36762 MW; BD91891649037700 CRC64;

Query Match 18.3%; Score 369; DB 16; Length 339;
Best Local Similarity 30.3%; Pred. No. 1.5e-20;
Matches 115; Conservative 56; Mismatches 162; Indels 46; Gaps 12;

```
OY 1 MKGLYYGTNDIRYSETPPEIKNPNDVKIKVSYCGICGTDLKEFTYSGGFVFPKQGT 60
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 1 MKAVGVRRKAYDLVTAE-VKKPVLKSKDEVLVKKRVGICGSDMHY-----HGT 48
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 61 KDKISGYELPLCPGHEFGTVEVSGVTSVKPGDRVAVEATSHCSDSRKYKOTVAQDLG 120
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 49 NPLAT---LPRVIGHEVTGOVEAVGANVQSLKPGDHVVEIPISYC-----G 91
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 121 LCMAGSGSPNCCASLFCGAGSAGFAEYVYGEDHWVKLPDSIPDDIGALVEPISYA 180
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 92 SCYACRKGPNVCARKLSVFGV-HEDGKREYTLPERQLHAVSKDLPWEAVMAEPTTIG 150
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 181 WHAVERARFOPGOTALVLGGPFGIATLILALOGHHAGKIVCSEPALIRROFAKELGAE-V 239
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 151 AQAVVRGQVREKGDYVLIGAGPIGIC-VLKMAKLAGAAMVMTDLNNERLAFKENGADAV 209
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 240 FDPSTCDANAVLKAMVPENEGFAAFDCSGVPQFTTTSIVATGPGIAYVNAVWGDHPI 299
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 210 VVVOAEHVAERVLE--WTGNEGANVYIDAVCLPETFALSIEAVSPAGHV-VLGFDERAA 266
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 300 GFMPSLTYOEKAYATGSMCYTWKDFOEYVKALEDGLISLQKARKMITGKV---HLKGYE 356
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 267 QISQLPITKKEVYITGSRLQT-NQPFKYVELLNG-----RLMHNGLVTHTFSSVDVH 318
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
OY 357 KGFKQLEIHKENNVKILVT 375
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
DB 319 HAFQFIKEHPDQVRKAVIT 337
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||
```

Search completed: May 2, 2003, 12:51:08
Job time : 1287 secs

GenCore version 5.1.4-PS-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 09:32:12 ; Search time 939 Seconds
(without alignments)
53.925 Million cell updates/sec

Title: US-10-020-674-2

Perfect score: 2013
Sequence: 1 MKGLLYGTNDIRSETVPE.....QLIEHKNNKILVTPEVS 380

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

_A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	567.5	28.2	363	22	AAU36386
2	538.5	26.8	370	23	ABB54213
3	519.5	25.8	345	23	AAE21046
4	491.5	24.4	354	23	ABP38529
5	455.5	22.6	352	22	AAU36382
6	417.5	20.7	291	22	AAE81823
7	360.5	17.9	357	15	AAE54779
8	358.5	17.8	348	23	ABB47897
9	357	17.7	350	22	AAE05988
10	355.5	17.7	360	22	ABB60809

11	352.5	17.5	354	22	AAE05985
12	352.5	17.5	360	22	ABE58718
13	349	17.3	316	22	ABB47473
14	348	17.3	343	23	ABB49510
15	346	17.2	368	23	AAE47254
16	342	17.0	358	22	AAU34594
17	342	17.0	363	12	AAE14181
18	337.5	16.8	343	22	AAE05994
19	337.5	16.8	350	22	AAE05966
20	336	16.7	343	23	AAU77018
21	334	16.6	352	22	AAE05987
22	331.5	16.5	348	22	AAE96816
23	329.5	16.4	348	21	AAE82330
24	328	16.3	352	21	AAE47840
25	328	16.3	365	21	AAE47839
26	328	16.3	370	21	AAE47838
27	327	16.2	301	22	AAE47474
28	327	16.2	346	21	AAE33242
29	327	16.2	359	21	AAE33241
30	327	16.2	364	21	AAE33240
31	326.5	16.2	356	21	AAE56748
32	326	16.2	341	22	AAE47471
33	323	16.0	346	21	AAE19409
34	323	16.0	359	21	AAE19408
35	323	16.0	364	21	AAE19407
36	319	15.8	343	22	AAU34209
37	319	15.8	354	22	AAU36871
38	315	15.6	347	22	AAU38180
39	314	15.6	352	22	AAE05993
40	308	15.3	341	22	AAU38240
41	307.5	15.3	346	22	AAU34613
42	307.5	15.3	346	22	AAE98373
43	306	15.2	342	22	AAU35365
44	303	15.1	336	22	ABB52481
45	300	14.9	317	21	AAE47846

ALIGNMENTS

RESULT 1	AAU36386	AAU36386 standard; Protein: 363 AA.
ID	AAU36386	
AC	AAU36386;	
DT	14-FEB-2002 (first entry)	
XX		
DE	Pseudomonas aeruginosa cellular proliferation protein #36.	
XX		
KW	Antisense: prokaryotic cellular proliferation protein; antibiotic; antibacterial; drug design.	
OS	Pseudomonas aeruginosa.	
XX		
PN	WO200170955-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	21-MAR-2001; 2001WO-US09180.	
XX		
PR	21-MAR-2000; 2000US-191078P.	
PR	23-MAY-2000; 2000US-206848P.	
PR	26-MAY-2000; 2000US-207727P.	
PR	23-OCT-2000; 2000US-242578P.	
PR	27-NOV-2000; 2000US-253625P.	
PR	22-DEC-2000; 2000US-257931P.	
PR	16-FEB-2001; 2001US-269308P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
XX		
PI	Hasselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ; Yamamoto RT, Xu HH;	

Bacillus haloduran
Drosophila melanog
C. max sorbitol de
Listeria monocytog
Arxula adeninivora
E. coli cellular p
Xylitol dehydrogen
Bacillus haloduran
Clostridium diffic
Idonic acid dehydr
Clostridium diffic
Putative P. abyssi
Pyrococcus horikos
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
T. aestivum sorbit
Arabidopsis thalia
Arabidopsis thalia
Human prostate can
Z. mays sorbitol d
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Staphylococcus aur
Salmonella typhi; c
Enterococcus avium
Salmonella typhi c
E. coli cellular p
Escherichia coli p
Haemophilus influe
Escherichia coli p
Arabidopsis thalia

```
XX WPI; 2001-611495/70.
DR N-PSDB; AAS54245.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS
XX Example 3; Seq ID No 11979; 511pp; English.
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 363 AA;
Query Match 28.2%; Score 567.5; DB 22; Length 363;
Best Local Similarity 36.1%; Pred. No. 7.5e-48;
Matches 136; Conservative 73; Mismatches 141; Indels 27; Gaps 11;
QY 1 MKGLVYGTNDIRYSEVPEPEIKNPNDYKIKYSGICGTDLKEFTYSGPVFFPKQGT 60
DB 11 MRAAVHGRDTRV-EDVPLAPPPGWOIRVHMGICGSDHE--YLAGPVFIPVE-A 66
QY 61 KDKISGYELPLCPGHEFSGTVEVSGVTSVKRQDRAVAEATSHCSDRSRKKTVAQDG 120
DB 67 PHLTGLKQCGILGHEFSGTVEVSGVTSVKRQDRAVAEATSHCSDRSRKKTVAQDG 112
QY 121 LCMACOSGSPNCASLSPFCGLGASGGAFAEYVYGGEDHMKLPDIPDDIGALVERISA 180
DB 113 TCYCHNGHYNICENLAFGLIMN-NGAFAYEVVWPANLLYALPAGPPEAGALIEPLANG 171
QY 181 WHAVERARFOPGOTATLVLGSGPIGLTIALOGHNAGKIVCSPPALIRQFAKELG-EY 239
DB 172 MRAVKAAGSLGONVYVAGTIGLSTIMCARAAGAOVIALEMSSARAKALEVASOV 231
QY 240 FDPSTCDDANAVLKAMVPNEGFHAFDCSGVPQFTTSIVATGPGSIANVAVMGDHP1 299
DB 232 LDPSSRC-DALGEIRALT-GGLGADVSFEICGNKHTAKLAIDARKKGKYLIVGIF-EES 288
QY 300 GFMPSLITVOEKYATGSMCTYVKDFQEVYKALEDLISDKARKMITGVHLKQVEKGF 359
DB 289 ENEFELVSTERKQLLGALYN-GEFADYIAFTADGRIDI---APLVTGRIGLEIEIVERG 344
QY 360 KOLIEHKENNVKILVTP 376
DB 345 EELVNNKEHNKVIIVSP 361
RESULT 2
ABB54213
ID ABB54213 standard; Protein; 370 AA.
XX
XX ABB54213;
XX
XX 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein butb.
DE
```

```
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX KW
XX OS Lactococcus lactis IL1403.
XX PN FR2807446-A1.
XX PD 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX Holotrine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species -
XX Claim 6; SEQ ID No 915; 2504pp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX nucleic acid sequence, particularly to identify Lactococcus lactis or
XX related species. The proteins of the invention are useful for the
XX biosynthesis or biodegradation of a composition of interest. The
XX invention helps research in lactic bacteria, particularly useful in the
XX production of yogurt and cheese.
XX Note: The sequence data for this patent is based on equivalent patent
XX WO200177334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 370 AA;
Query Match 26.8%; Score 538.5; DB 23; Length 370;
Best Local Similarity 37.3%; Pred. No. 6.2e-45;
Matches 144; Conservative 53; Mismatches 144; Indels 45; Gaps 14;
QY 1 MKGLVYGTNDIRYSEVPEPEIKNPNDYKIKYSGICGTDLKEFTYSGPVFFPKQGT 60
DB 15 MRAARFYDGDRIIDE-INELPYK-AGQYIDVAMGICGTDLHEFL--DGEIFCSAEH 70
QY 61 KDKISGYELPLCPGHEFSGTVEVSGVTSVKRQDRAVAE--ATSHCSDRSRKKTVAOD 118
DB 71 PNLITGEVPPVTLGHMSGVVNEIGEGVGLKVGDHVVEPYIVPEGTDTs----- 121
QY 119 LGLCMACOSGSPNCASLSPFCGLGASGGAFAEYVYGGEDHMKLPDIPDDIGALVERIS 178
DB 122 -----ETGHNLSGSGNFILGNGGLAEKISVDENWVHKIPDNLDEALIEPIS 174
QY 179 VAMHAVERARFOPGOTATLVLGSGPIGL-ATIALOGHNAGKIVCSPPALIRQFAKELG 236
DB 175 VGHHAVERANLEKSTVLVVGAGPIGLLAAVAKQGH---IYIISSEGLRRKKAQE-- 229
QY 237 AEV----FDPSTCDDANAVLKAMVP-NEGFHAFDCSGVPQFTTSIVATGPGSIANV 290
DB 230 AQVADYFENPIEDD-----IQAKVHEINKEGVDAEFECTSVOPGPDACDAIRMGDTYV 284
QY 291 VAVMGDHP1GFMPSLITVOEKYATGSMCTYVKDFQEVYKALEDLISDKARKMITGVH 350
DB 285 VAIWG-KRASVDMAKIVIKKANLGLTIAVN-NTHKRTIDLVSTGKIKLD---QFTIAKIG 339
QY 351 LKQVEKGFKOLIEHKENNVKILVTP 376
DB 340 LDDLIDKGFDTLIIHNETAIVKILVSP 365
RESULT 3
AAE21046
```

ID AAE21046 standard; Protein; 345 AA.
 AC AAE21046;
 XX
 XX
 DT 01-JUL-2002 (first entry)
 XX
 XX
 DE Human drug metabolizing enzyme (DME-4) protein.
 XX
 KM Human: drug metabolizing enzyme; cell proliferative disorder; metabolic;
 KM autoimmune; inflammatory; developmental; gastrointestinal; hypergonadal;
 KM pancreatic; endocrine; eye; dermatitis; Addison's disease; antihypertensive;
 KM acquired immunodeficiency syndrome; AIDS; glomerulonephritis; anorexia;
 KM diabetes; atherosclerosis; adult respiratory distress syndrome; anaemia;
 KM Grave's disease; thyroiditis; Crohn's disease; infection; anticoagulant;
 KM systemic lupus erythematosus; cirrhosis; psoriasis; epilepsy; gastritis;
 KM cataract; hypopituitarism; cancer; rheumatoid arthritis; conjunctivitis;
 KM cystic fibrosis; peptic ulcer; Wilson's disease; hepatitis; anticholinergic;
 KM allergy; diarrhoea; thrombosis; obesity; immunostimulant; tranquilizer;
 KM infertility; vulvovaginitis; anticonvulsant; gynaecological; laxative; goitre;
 KM neotropic; jaundice; trauma; asthma; DME-4; enzyme.
 XX
 OS Homo sapiens.
 XX
 OS
 PN W0200212467-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 03-AUG-2001; 2001MO-US24382.
 XX
 PR 04-AUG-2000; 2000US-223055P.
 PR 11-AUG-2000; 2000US-224728P.
 PR 18-AUG-2000; 2000US-226440P.
 PR 24-AUG-2000; 2000US-228067P.
 PR 31-AUG-2000; 2000US-230063P.
 PR 13-SEP-2000; 2000US-232244P.
 PR 20-SEP-2000; 2000US-234269P.
 XX
 PA (INCYTE) INCYTE GENOMICS INC.
 XX
 PI Baughn MR, Bruns CM, Das D, Delegeane AM, Ding L, Elliot VS;
 PI Gandhi AR, Griffin JA, Hafalla AJA, Khan FA, Lal P, Lee S;
 PI Lu DMH, Lu Y, Patterson C, Ramkumar J, Ring HZ, Sanjanwala MS;
 PI Tang YT, Thangavelu K, Thornton M, Tribley CM, Walla NK;
 PI Warren BA, Yang J, Yao MG, Yue H;
 XX
 DR WPI: 2002-206331/26.
 DR N-PSDB; AAD33483.
 XX
 PT New human drug metabolizing enzyme polypeptide and polynucleotide
 PT useful for diagnosing, treating and preventing cell proliferative,
 PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal
 PT disorders -
 XX
 PS Claim 48: Page 149-150; 179pp; English.
 XX
 CC The invention relates to an isolated human drug metabolizing enzyme (DME)
 CC polypeptide or a biologically active or immunogenic fragment of DME. DME
 CC is useful for diagnosis, treatment and prevention of cell proliferative,
 CC autoimmune/inflammatory, developmental, endocrine, eye, metabolic and
 CC gastrointestinal disorders including live disorders. Autoimmune/
 CC inflammatory disorders include acquired immunodeficiency syndrome (AIDS),
 CC adult respiratory distress syndrome, Addison's disease, atherosclerosis,
 CC allergies, anaemia, asthma, autoimmune haemolytic anaemia, autoimmune
 CC thyroiditis, Crohn's disease, atopic dermatitis, diabetes mellitus,
 CC glomerulonephritis, rheumatoid arthritis, systemic lupus erythematosus,
 CC ulcerative colitis, uveitis, viral, bacterial, protozoal, parasitic,
 CC fungal, helminthic infections and trauma. Cell proliferative disorders
 CC include cancer, arteriosclerosis, cirrhosis and psoriasis; developmental
 CC disorders include epilepsy and cataract; and endocrine disorders include
 CC disorders of hypochalams/pituitary, disorders associated with
 CC hypopituitarism, including diabetes insipidus, hypogonadism, disorders
 CC associated with hypothyroidism including goitre, Grave's disease,
 CC pancreatic disorders such as diabetes mellitus, disorders associated with

CC adrenals, disorders associated with gonadal steroid hormones such as
 CC endometriosis, infertility, hypergonadal disorders and gynecomastia.
 CC Disorders of the eye include conjunctivitis and macular degeneration and
 CC metabolic disorders include diabetes, cystic fibrosis, obesity and
 CC hypocalcaemia. Gastrointestinal disorders include gastritis, peptic
 CC ulcer, hepatitis, constipation, diarrhoea, jaundice, Wilson's disease,
 CC thrombosis and hepatic tumours. DME gene is useful in gene therapy. The
 CC present sequence is human DME-4 protein.
 XX
 SQ Sequence 345 AA:
 XX
 Query Match 25.8%; Score 519.5; DB 23; Length 345;
 Best Local Similarity 35.4%; Pred. No. 4.4e-43;
 Matches 134; Conservative 62; Mismatches 142; Indels 41; Gaps 12;
 QY 1 MKGLLYGTNDIRSEVPEPEINPDVIRKYSYICIGTDLKEFYSSGVPFPFGQGT 60
 DB 1 MKAAMWYGQKDVREER--EPKELQDNEVYKVMAGICGTDLHE--YLEGPFISTE-K 55
 QY 61 KDKISGYELPLCPGHEFSGTVEVSGSVTSVKGPDRAVAVTSHCSDRSRKQTVADLQ 120
 DB 56 PDPFLGQKAPVTLGHFAGVVEETGSGVTKFNKDRVAVPT--VSKREKENIDLYD-- 111
 QY 121 LCMAGQSGSPNCASLSPGGLGASGGFAEYVYGGEDHMKLPDIPDIDIGALVEPISVA 180
 DB 112 -----GYSFICL-GSDGGFAETNPAPENYKLPDNPVSDKCALVEPTAVA 156
 QY 181 WHAVERARFGQRYALVGGGPIGLATITLALQGHAKIKYCEPALLRRQFAKLGA-EV 239
 DB 157 VQATKEGEVLFGDVAIFGAGPIGLTLVVAKAGASKIFVFDISERLKAKALGATHA 216
 QY 240 FDPSTCDANAVLKAMPENE---GFHAFFDCSGVQFTFTTSIVATGPGSGLAVNVWVG 295
 DB 217 INSGKTPPVYI-----NEYTENGVDVSEVAGVAPTLKSSLDVTAKAGTYIVSIFG 269
 QY 296 DHPIGFMPMSLYOEKAYATGSMCTYKDFQEVYKALBDGLISDKARKMITGKYLKDG 355
 DB 270 -HPLEWNPQDLTNTGVKLTSTIAYTPTTFQOTDLINENGLNV---KDVYDELELNIV 325
 QY 356 EKGFKOLEIKENNVKLV 374
 DB 326 ESGFEQLVNDK-SQAKILI 343
 XX
 RESULT 4
 ABP38529
 ID ABP38529 standard; Protein; 354 AA.
 XX
 AC ABP38529;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3374.
 XX
 KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KM antibacterial; gene therapy.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 PA (GENO-) GENOME THERAPEUTICS CORP.
 XX
 PI Doucette-Stamm LA, Bush D;
 XX
 DR WPI: 2002-381255/41.

DR N-PSDB; ABN91074.
XX Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 3374; 267pp; English.
XX
XX ABN0538 to ABN9374 represent *Staphylococcus epidermidis* open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The *S. epidermidis* sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly *S. epidermidis* infections. The sequences can be used to
CC screen for compounds able to interfere with the *S. epidermidis* life
CC cycle or inhibit *S. epidermidis* infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 354 AA:
Query Match 24.4%; Score 491.5; DB 23; Length 354;
Best Local Similarity 34.9%; Pred. No. 2.9e-40;
Matches 132; Conservative 59; Mismatches 154; Indels 33; Gaps 11;
QY 1 MGGLTYGTNDIRYSETPPEIKNDYKIRVYSGICGTDLKERTYSGGPFPPKQGT 60
DB 5 MKAAWYGGKDVAVEDR--EPKAIKNEVQVAVSMAGICGTDLHE--YLEGPIFT-STQO 59
QY 61 KDKISGYELPLCPGHEFSGTVEVSGSVTVKRGDRAVEATSHCSRSRYKDYVAQDGL 120
DB 60 PPLILGQTAPVTLGHEFSGVENVKNSRFKKGDRVYVNP--VSKREKPEVNDLYD-- 115
QY 121 LCMACOSSPNCASISFSGIGASGSGFAEYVYGGEDHWKLPDSIPDDIGALVEPISVA 180
DB 116 -----GYSFISGL-GSDGAFEFNAFTNAPNTNYHLDPNVSAREGALVEPTAVA 160
QY 181 WHAVERARFQPGOTALVLGCGPIGLATILALOGHHAGKIVCSEPALIRROFAKELGA-EV 239
DB 161 VQAVKRGELLFQGTVAVFAGAPIGLITLIVAAKAAGASKIFVPLDSERLAKKSVGATHV 220
QY 240 FDPSTCDDANAVLKAVNPENEGHAFDCSGVPQFTTSIVATGSGIAVNVAVMGDHP 299
DB 221 YNSGNDVPQVTAYE--HFDNGVDVSFEVAGVITLQOSIEVTRPRGTAVIYSIFG-HFV 276
QY 300 GFMPMSLTYQEKYATGSMGYTKVDFOEVYKALEDGLISDKARKMTTGKVGKGEKGF 359
DB 277 EFNPLQNNKGYKLTITTYPTTFQOTIDLIANGSLNV---KDVVTDQIEIDNIVESGF 333
QY 360 KQIIEKENNVKILVTPN 377
DB 334 NQLVNDK-SQAKTIVRLN 350
RESULT 5
AAU36382
ID AAU36382 standard; Protein; 352 AA.
AC AAU36382;
XX
XX 14-FEB-2002 (first entry)
XX
XX *Pseudomonas aeruginosa* cellular proliferation protein #372.
DE
XX
XX Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
OS
XX *Pseudomonas aeruginosa*.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX

PF 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207127P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54241.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 11975; 511pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 352 AA:
Query Match 22.6%; Score 455.5; DB 22; Length 352;
Best Local Similarity 33.5%; Pred. No. 1.2e-36;
Matches 127; Conservative 55; Mismatches 162; Indels 35; Gaps 12;
QY 1 MGGLTYGTNDIRYSETPPEIKNDYKIRVYSGICGTDLKERTYSGGPFPPKQGT 60
DB 1 MKALRWHAARDLRLSEL--ERQAPRGEVLELVAVAGICGSDLHE--YQSGHSP-QAE 55
QY 61 KDKISGYELPLCPGHEFSGTVEVSGSVTVKRGDRAVEATSHCSRSRYKDYVAQDGL 120
DB 56 AHPISGCRAPLTLGHEFCVVAALGEGVGPRIQDRAVEAPEYRC-----G 101
QY 121 LCMACOSSPNCASISFSGIGASGSGFAEYVYGGEDHWKLPDSIPDDIGALVEPISVA 180
DB 102 ECRYCEGRYNLCESMGFTGLMG-DGFAERARVPAYMLHRLPDVAGFQAALVEPAAYA 160
QY 181 WHAVERARFQPGOTALVLGCGPIGLATILALOGHHAGKIVCSEPALIRROFAKELGA-EV 239
DB 161 LHALRRSSILAPGRCVAPFLGPIGLLVLARLGRJEDIAADVSPERLALGEGASRA 220
QY 240 FDPSTCDDANAVLKAVNPENEGHAFDCSGVPQFTTSIVATGSGIAVNVAVMGDHP 299
DB 221 LPARGDGTA-----ARLRGALDCAFEAAGSASIDALASIRKGGELVYLNGE--V 273
QY 300 GFMPMSLTYQEKYATGSMGYTKVD--FOEVYKALEDGLISDKARKMTTGKVGKGEKGF 358
DB 274 RLDAPDLVNRRLILGSGVY--RDAVPELIALLAGRLDLARA---VTRSVPLEQAVENG 328

Db 20 EQRPIPIKDPHYKALIKATIGCGSDI-HYRSGG-----IGKYLKADMYL 66
 QY 74 GHEPSTVEVSGSVTKGDRVAVEATSHCSRSRYKDTVAODLGLMACGSGSNCC 133
 Db 67 GHESGQVVEGDAVTVKVGDRVAIEP-----GVPSRISDET-----KEGRYNLC 112
 QY 134 ASISFCGLGASGGFAEYVYVGGDHMKLPDIPDDIGALVEPISVAMHAVERARFQPGQ 193
 Db 113 PHMAFATPIPDGTLYKTYLSPEDFLVKLPESGYEGACVEPLSVGHKNKLAVGFRT 172
 QY 194 TALVGGGPIGLATITIALOGHNAGKIVCSSEPALIRQPAKELGA-EVFDPS--TCDDANA 250
 Db 173 KVVVFGAPVGLLTGAVARFAGATDIFVDFDNKLRANDFGATNTFNSSQFSTDKAOD 232
 QY 251 VLKAMPENEGFHA--AFDCSGVQPTFTSIVATGPGSIANVAVMGDHPDIGEMPSLTLY 308
 Db 233 LAGGVKLLGNGNADVVEFGSGADVCIDAAVKTKTGVGTWVQGM-GKNTNFPPIAEVSG 291
 QY 309 QERYANGSMCYTVKDFQEVVKALEDGLISLDKARKMITGKVLKDGVEKGFOLIEHKN 368
 Db 292 KEMKLIGCFRYSFGDYRDVAVNLVATGKVN---KPLTHKFPEDAA-KAYDYNIAHGGE 347
 QY 369 NVKILY 374.
 Db 348 VKRTII 353

RESULT 8
 ABB47897
 ID ABB47897 standard; Protein; 348 AA.
 AC ABB47897;
 XX
 DT 05-FEB-2002 (first entry)
 DE Listeria monocytogenes protein #601.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 OS Listeria monocytogenes.
 PN WO200177335-A2.
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Ruhnok C, Esini H, Dehoux P,
 PI Dusserget O, Chetouani F, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides
 XX
 PS Claim 6; SEQ ID No 602; 192pp; French.
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic

CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 348 AA:

Query Match 17.8%; Score 358.5; Db 23; Length 348;
 Best Local Similarity 28.3%; Pred. No. 5,7e-27;
 Matches 110; Conservative 65; Mismatches 159; Indels 55; Gaps 11;

QY 1 MGILYYGTNDIRYSEVPEPEIKNPNDYKIKVSTGICGTDLKEPTYGGPVFPKQGT 60
 Db 1 MKALKYGRKDLRYEE-ADMPTEEPDDVILKVTYIGCGSDISRYT----- 46
 QY 61 KDRISGYELPLCPGHEPSTVEVSGSVTSVKRGDRVAVEATSHCSRSRYKDTVAQDLG 120
 Db 47 --KLGPYVGMWGHFSGEIVSEVTDIEIGDAACAPALYC-----G 90
 QY 121 LCMACGSGSPNCASLSPFCGLGASGFAEYVYVGGDHMKLPDIPDDIGALVEPISVA 180
 Db 91 ECEYCKRGEFARCKRLTVIG-ARHPGAVAEYIKLPANVVKIPNELDYEAALVPSAVV 149
 QY 181 MHAVERARFQGTALVLDGGPIGLATITIALOGHNAGKIVCSSEPALIRQPAKELGAEPV 240
 Db 150 VHEFYHTKLDAGDDVVVSGNIGLAIQMAKFGARVAIDVDKRLALAKEVGADVV 209
 QY 241 DPSTCDANAVLKAMPENEGFHA--AFDCSGVQPTFTSIVATGPGSIANVAVMGDHP 298
 Db 210 INSLKEDP---LEVVAHHDGLNADLVNRAAGSPIT-SQGVRAVAKGGV---VFLGIP 262
 QY 299 IGFMPMSLTYYQEKYATGSMC-----YTVKDFQEVVKALEDGLISLDKARKMITG 347
 Db 263 YADVTLIERFYEKIVASELTWMSGWNAISAPFGKEMQTTIHFLANKQINIE---PMITH 319
 QY 348 KYHLKDGVEKGFOLIEHKNKILVTP 376
 Db 320 RLSLAGPEV-FERYIERNEFFGKVLFFP 347

RESULT 9
 AAE05988
 ID AAE05988 standard; Protein; 350 AA.
 AC AAE05988;
 XX
 DT 24-SEP-2001 (first entry).
 DE Clostridium difficile homologue protein #3.
 XX
 KW Five-carbon sugar; aldo-sugar; keto-sugar; sugar alcohol; fermentation;
 KW pentose phosphate pathway; xylitol; D-arabitol; D-arabinose; D-lyxose;
 KW ribitol; D-ribose; D-ribulose; D-xylulose; D-xylulose; microbial host;
 KW arabitol phosphate dehydrogenase; APDH; xylitol phosphate dehydrogenase;
 KW XPDH.
 OS Clostridium difficile.
 OS
 PN WO200153306-A2.
 PD 26-JUL-2001.
 XX
 PR 22-JAN-2001; 2001WO-FI00051.

XX 24-SEP-2001 (first entry)
 XX
 XX Bacillus halodurans sorbitol dehydrogenase homologue protein #1.
 DE
 XX
 XX Five-carbon sugar; aldo-sugar; keto-sugar; sugar alcohol; fermentation;
 KM pentose phosphate pathway; xylitol; D-arabitol; D-arabinose; D-lyxose;
 KM ribitol; D-ribose; D-xylulose; D-xylitol; microbial host;
 KM arabitol phosphate dehydrogenase; APDH; xylitol phosphate dehydrogenase;
 KM xPDH; sorbitol dehydrogenase.
 XX
 XX Bacillus halodurans.
 OS
 XX WO200153306-A2.
 PN
 XX
 XX 26-JUL-2001.
 PD
 XX
 XX 22-JAN-2001; 2001WO-FI00051.
 XX
 XX 21-JAN-2000; 2000US-0488581.
 XX
 XX (XYRO-) XYROFIN OY.
 XX
 XX Miasnikov A, Ojamo H, Povelainen M, Gros H, Toivari M, Richard P,
 PI Ruomonen L, Koivuranta K, Londenborough J, Aristidou A;
 PI Penttilae M, Plazenet-Meunt C, Deutscher J;
 XX
 XX WPI: 2001-465360/50.
 DR
 XX
 XX Isolated polynucleotide, used to transform bacterial or yeast hosts
 PT which can then be used in the production of sugars and sugar alcohols,
 PT encodes xylitol phosphate dehydrogenase -
 XX
 XX Claim 14; Page 197-198; 205pp; English.
 XX
 XX The present invention relates to the methods for manufacturing
 CC five-carbon aldo- and keto-sugars and sugar alcohols by fermentation in
 CC recombinant hosts. The recombinant hosts of the invention have been
 CC engineered to enhance the production of the pentose phosphate pathway
 CC intermediates, or the production of one or more of xylitol, D-arabitol,
 CC D-arabinose, D-lyxose, ribitol, D-ribose, D-xylulose, D-xylitol and/or
 CC D-xylulose. Arabitol phosphate dehydrogenase (APDH) is used in a
 CC microbial host cell to produce recombinant arabitol. Xylitol phosphate
 CC dehydrogenase (XPDH) and arabitol phosphate dehydrogenase are used in a
 CC microbial host cell to produce recombinant xylitol. The present sequence
 CC is Bacillus halodurans sorbitol dehydrogenase protein which is
 CC homologous to Lactobacillus rhamnosus xPDH protein.
 XX
 XX Sequence 354 AA:
 SQ
 Query Match 17.5%; Score 352.5; DB 22; Length 354;
 Best Local Similarity 27.6%; Pred. No. 2,3e-26;
 Matches 108; Conservative 56; Mismatches 176; Indels 51; Gaps 9;
 OY 1 MGGLLYGTNDIRSEYVEPEIKNPNDVKIKVSYCGTGLDKEFTYSGGVPEPKQGT 60
 DB 1 KMALMLYGIQIDRFREET-PAPSIHEHDDIIIKVKAIVGICGSLSTR----- 45
 OY 61 KDKISGYELPLCPGHEFGTVEVSGVTSVRPGDRVAVEARSHCSDSRKYDTYAQDLG 120
 DB 46 -KLLGPYPVPGMTFGHEFGEVVKIGRSVTFSGIDRVAACTPYTC-----G 90
 OY 121 ICMACQSGSPNCCASLSTFCGLGASGFAEYVYGEDHNVKLPDSIPDDIGALVEPISVA 180
 DB 91 QCRVQGLDEPTFCERLSTYIG-ARHGAAYAEYKLPKAKHVIPLPNVNVNDEALIPASVY 149
 OY 181 WHAVERARPGQGTALVLGGPGTIGATLALOGHAGKIVCSSEPALIRROFKAELGAEEV 240
 DB 150 AHGFRTIKIKPGASVAIVGSGISGLAVQMAKIFGATVFALIDDEOKLVANVANGADVL 209
 OY 241 DPSTDDNNVILKAMVPENEGHAFFDCSGVPQFTTTSIVATPGSGIANNVAVMGDHPIG 300
 DB 210 ISSIQRPAAHKQILEYT-NGIGVAVVESAGTPTSTAQVFALPKKGGEVVFLGI-----PYA 264

OY 301 FPMSTLYQEKYATG-----SMCYTWKDFQEVVKALEDGLISDKARKMITGKV 349
 DB 265 DVQIEKFEYFEKIVRNELHYGSMNALSSPPEKEMTTIHHSSQGLNV---APMISTYL 321
 OY 350 HLKDGVEKGFQKQILIEHKNVKKILVTPNEVS 380
 DB 322 PLAKGPET-FOQJAKGELKPTVLEFPEKLS 351
 RESULT 12
 ABB58718
 ID ABB58718 standard; Protein; 360 AA.
 AC
 XX ABB58718;
 AC
 XX
 XX 26-MAR-2002 (first entry)
 DT
 XX
 XX Drosophila melanogaster polypeptide SEQ ID NO 2946.
 DE
 XX
 XX Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 KM
 XX Drosophila melanogaster.
 OS
 XX WO200171042-A2.
 PN
 XX
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX
 XX WPI: 2001-656860/75.
 DR
 XX N-PSDB; ABL02821.
 DR
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more;
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure: SEQ ID NO 2946; 21pp + Sequence Listing; English.
 PS
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPRO
 CC at ftp.wipro.int/pub/published_pcf_sequences.
 CC
 XX
 XX Sequence 360 AA:
 SQ
 Query Match 17.5%; Score 352.5; DB 22; Length 360;
 Best Local Similarity 28.3%; Pred. No. 2,4e-26;
 Matches 106; Conservative 63; Mismatches 161; Indels 45; Gaps 11;
 OY 7 YGTNDIRSEYVEPEIKNPNDVKIKVSYCGTGLDKEFTYSGGVPEPKQGTDKISG 66
 DB 11 GHIEDMR-L-EQRPIDEAD-DEVILAMDSVIGCGDVHYLAHG-----RIGD 55
 OY 67 YEL--PLCPGHEFGTVEVSGVTSVRPGDRVAVEARSHCSDSRKYDTYAQDLGLMA 124
 DB 56 FVLTRPKLIIHGHSAGVAVKLAKKVTTLKGVRALEPQVPCR-----CDH 101

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:05:52 ; Search time 622 Seconds
(without alignments)
58,732 Million cell updates/sec

Title: US-10-020-674-2

Sequence: 1 MKGLLYGTNDIRSEVPE.....QLEHKENKILVTPNEVS 380

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	923.5	45.9	382	2	FS1962
2	912	45.3	417	2	FS1961
3	571	28.4	348	2	EB4143
4	567.5	28.2	363	2	CS3128
5	551.5	27.4	346	1	H69789
6	538.5	26.8	370	2	AB6737
7	537	26.7	354	2	EB1178
8	534	26.5	354	2	H81925
9	455.5	22.6	352	2	AB3139
10	422.5	21.0	357	2	AB3139
11	422.5	20.3	397	2	AB3052
12	408.5	20.3	353	1	AA5052
13	386.5	19.2	347	2	T23889
14	370.5	18.4	348	1	S32484
15	370	18.4	395	2	B72381
16	369	18.3	339	1	F69852
17	368.5	18.3	347	2	T23890
18	365.5	18.2	357	1	SE6811
19	364.5	18.1	348	2	AB1496
20	362	18.0	360	1	S35981
21	360.5	17.9	357	1	S55941
22	358.5	17.8	348	2	AC1138
23	352.5	17.5	354	2	CS3673
24	348	17.3	343	2	AF1407
25	348	17.3	343	2	AF1783
26	346	17.2	347	2	AH2749
27	346	17.2	371	2	G97530
28	345	17.1	358	2	E90939
29	345	17.1	358	2	AB5788

30	343	17.0	343	2	E95295	1dnd L-Idonate 5-d
31	342	17.0	358	1	H64937	probable alcohol d
32	342	17.0	363	1	S13529	D-xylose reducta
33	339	16.8	343	2	F98227	L-Idonate 5-dehydr
34	339	16.8	343	2	AB3059	zinc-binding dehyd
35	338	16.8	371	2	AB0407	probable zinc-bind
36	337.5	16.8	343	2	EB3673	sorbitol dehydroge
37	337.5	16.8	389	2	AF1918	alcohol dehydrogen
38	337.5	16.8	399	2	S16132	L-Iditol 2-dehydro
39	337	16.7	347	1	F64937	probable L-Iditol
40	336.5	16.7	354	1	S10065	L-Iditol 2-dehydro
41	336	16.7	357	1	A54674	L-Iditol 2-dehydro
42	333.5	16.6	356	1	S64902	probable sugar red
43	331.5	16.5	348	2	G75049	L-threonine 3-dehy
44	330	16.4	347	2	C90939	probable oxidoredu
45	330	16.4	347	2	G85787	probable oxidoredu

ALIGNMENTS

RESULT 1

FS1962

FUN49 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YAL060w

C:Species: Saccharomyces cerevisiae

C>Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: FS1962

R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; submitted to the EMBL Data Library, August 1994

A:Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.

A:Reference number: S51956

A:Accession: S51962

A:Molecule type: DNA

A:Residues: 1-382 <BUS>

A:Cross-references: EMBL:U12980; NID:91326053; PID:9595526; GSPDB:GN00001; MIPS:YAL0

C:Genetics:

A:Gene: SGD:BDH1; FUN49; MIPS:YAL060w

A:Cross-references: SGD:S0000056

A:Map position: 1L

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology <LADH>

F:24-363/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match	45.9%	Score 923.5	DB 2	Length 382
Best Local Similarity	48.7%	Pred. No. 2.9e-63		
Matches 184	Conservative 65	Mismatches 126	Indels 3	Gaps 2
1	MKGLLYGTNDIRSEVPEPIKPNNDVKKIVSYCGICGTDLKEFTYSGGPVFPKQCT	60		
1	MAALAVFPRKGDHFNIDIPREIOTRDEVIIDVSMCGISGDLHE--YLDGPIFMRKDE	58		
61	KDKIGYELPLCPGHEFGSTVVEGCVTSVAPGRVAVENATSHSDDSRYVDYAOGLG	120		
59	CHKLNNALPLMGHEMSGIVSKVGPVKYVGVGVVDAASSCADLHCWHSFYSK	118		
121	ICMACOSGSPNCASLSFGGASGGAFFVYGEHDHVKLPDIPDIDIGALVPISYA	180		
119	PCDAOGRSENCTHAGFVGLVIGSGFRAEQVYVSGHNIIPKPEIPDLVALVLEPST	178		
181	WHAVERARPOGOTALVLGGPGLATILALOGHHAGKIVCEPALIRROFAKEIGAEVF	240		
179	WHAVERISGRKSSALVIGAGPIGLCTILVLMGMSKIVSEIARREMAKILGVEVF	238		
241	DPSTCDAN-AVLKAMPENEGFNAFDGSSVPOTFTTSIVATGSGIVANVAVGCDHI	299		
239	NPSKHGHSKIELRGLTKSHDEFDYSYDSCGIQVFFESLKLTKRGATNLAIVGPKV	298		
300	GFMPSLTQEKYATGSMCTYKQPOEVYKALEDLGISLDRKMKITGVHLKDEVEKF	359		
299	PROPDVTLQEKVMTGSIYVVEAEVYRAIHNGDIAMDEKQILITGQRIEDWKEGF	358		
360	KOLIEHKENKILVTPN 377			

Db 359 QELMDHKNVKKILTPN 376

RESULT 2

S51961

FUN50 protein - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: protein YAL061w

C:Species: *Saccharomyces cerevisiae*

C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S51961

R:Bussey, H.; Kadoy, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Oneil

Submitted to the EMBL Data Library, August 1994

A:Description: The sequence of chromosome 1 of *Saccharomyces cerevisiae*.

A:Reference number: S51956

A:Accession: S51961

A:Molecule type: DNA

A:Residues: 1-417 <BUS>

A:Cross-references: EMBL:U12980; NID:g1326053; PID:g959525; GSPDB:GN00001; MIPS:YAL061w

C:Genetics:

A:Gene: FUN50; MIPS:YAL061w

A:Cross-references: SGD:S0000057

A:Map position: 1L

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology <LADH>

F:24-364/Domain: long-chain alcohol dehydrogenase homology <LADH>

Query Match 45.3%; Score 912; DB 2; Length 417;

Best Local Similarity 45.9%; Pred. No. 2.5e-62;

Matches 174; Conservative 71; Mismatches 130; Indels 4; Gaps 2;

QY 1 MKGLIYGTNDIRYSEVPEPEIKNPNDVYKIKSYGICGTDIKFTYSGGPVFPKQGT 60

DB 1 MRLAIFGKGNIRPTNMLKEPHIVAPDELVIDEMCGICGTDLHE--YTDGPIFFPEDGH 58

QY 61 KDRISGEYELPCGHEHSGTVVSGVTSVKRPGDRVAVATSHCSDRSKYKDTVAQDLG 120

DB 59 THEISHNPILPQANGHEAGVLEVGKVKLKGDKYVVEPTCTCRDRIYWPISPNVDKE 118

QY 121 LCMACQSGSPNCCASLFCGLGASGGAFAEYVYGGEDHMYKLPDIPDICALVEPISVA 180

DB 119 WCAACKRGYNYISYGLCLGAGVSGGFAERVYNNESHCKYVDPFLVLAALQPLAVC 178

QY 181 WHAVERARPOGOTLVIGGGPIGLATLIALOGHNAGKIVCSPALIRROFAKELGAEV 240

DB 179 WHAIRCERAGSTALIGAPIGLGLTIALNAGCKDKVIDSPAVRRELAKMKARVY 238

QY 241 DPG--TCDDANAVLKAMVPENEGFAHAFDCSGVPQFTTSIVATGPGSGLAVNAVWGDHP 298

DB 239 DPTAHAKESIDYLRISADGGDFDYTFDCSGLEVTLNALQCLTFRGTAVNLAAMGHHK 298

QY 299 IGFPMPLTYOEKATGSCYATPKDQEVYKALDELISLDRARKKITGKVLKQGVK 358

DB 299 IOFSPMDITLHERKYTGSMCYTHHDEAVILEALEGRIDIRARHMITGRVNIEDLDGA 358

QY 359 FKOLIEKENNVKILVTPN 377

DB 359 IMKLINKESTIKILTPN 377

RESULT 3

EB4143

L-iditol 2-dehydrogenase BH3949 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: EB4143

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: EB4143

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-348 <STO>

A:Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BA07668.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3949

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 28.4%; Score 571; DB 2; Length 348;

Best Local Similarity 37.7%; Pred. No. 3e-36;

Matches 141; Conservative 65; Mismatches 134; Indels 34; Gaps 13;

QY 1 MKGLIYGTNDIRYSEVPEPEIKNPNDVYKIKSYGICGTDIKFTYSGGPVFPKQGT 60

DB 1 MQLRHWGADVADVADIAEPATLK--GKVKIKKMGICGSDLHE--YAGPIFIP-QNS 55

QY 61 KDRISGEYELPCGHEHSGTVVSGVTSVKRPGDRVAVATSHCSDRSKYKDTVAQDLG 120

DB 56 PHSUTEKAPIYVGHESGTVVSGVTSVKRPGDRVAVATSHCSDRSKYKDTVAQDLG 101

QY 121 LCMACQSGSPNCCASLFCGLGASGGAFAEYVYGGEDHMYKLPDIPDICALVEPISV 179

DB 102 TCTACRQKGNLCEOLGFLGAGGGFSEYTV--DEHMYKIPDYVSPGALVEPAAV 160

QY 180 AMHABARPOGOTLVIGGGPIGLATLIALOGHNAGKIVCSPALIRROFAKELGAEV 239

DB 161 ALYAVROSQFNVDQAVVEGTGPIGLVLEALKASASKIYAVELSTERORAEOLGATA 220

QY 240 FDPSTDDANAVLKAMVPENEGFAHAFDCSGVPQFTTSIVATGPGSGLAVNAVWGDHP 299

DB 221 LNPA---EVNVAEIRLDGADVSEVTVLQALSTKLGQTMIVSIF-EKRA 276

QY 300 GFMPLTYOEKATGSCYATPKDQEVYKALDELISLDRARKKITGKVLKQGVK 358

DB 277 SFQPNLVYKRRIRNIGLIG--RDIPAVISLMNGYFS---AETLVTRIGLHDVBOG 331

QY 359 FKOLIEKENNVK 371

DB 332 FEALM--KTNCKIK 343

RESULT 4

CB3128

2,3-butanediol dehydrogenase PA4153 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: CB3128

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; L

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: CB3128

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <STO>

A:Cross-references: GB:AE004831; GB:AE004091; NID:g9950347; PIDN:AG07540.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 28.2%; Score 567.5; DB 2; Length 363;

Best Local Similarity 36.1%; Pred. No. 5.8e-36;

Matches 136; Conservative 73; Mismatches 141; Indels 27; Gaps 11;

QY 1 MKGLIYGTNDIRYSEVPEPEIKNPNDVYKIKSYGICGTDIKFTYSGGPVFPKQGT 60

DB 11 MAAVAWHHRHDIR-EDVPLPAEPFGVQIRVHMCICGSLHE--YLAGVFLPVE-A 66

QY 61 KDRISGEYELPCGHEHSGTVVSGVTSVKRPGDRVAVATSHCSDRSKYKDTVAQDLG 120

DB 67 PHRLGLKDDCGLGHEFGEIYRLGNGYTGRAVGAANAADCCQH-----G 112

QY 121 LCMACQSGSPNCCASLFCGLGASGGAFAEYVYGGEDHMYKLPDIPDICALVEPISVA 180

E81178
Probable alcohol dehydrogenase (EC 1.1.1.-) NMB0604 [similarity] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: E81178
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Hatt, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
H. Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappolt, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:2015755; PMID:10710307
A:Accession: E81178
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <TEXT>
A:Cross-references: GB:AE002416; GB:AE002098; NID:g7225832; PIDN:AAF41031.1; PID:g722583
A:Experimental source: serogroup B, strain MC58
A:Genetics: NMB0604
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase

Query Match 26.7%; Score 537; DB 2; Length 354;
Best Local Similarity 36.0%; Pred. No. 1.2e-33;
Matches 140; Conservative 54; Mismatches 147; Indels 48; Gaps 12;

1 MGGLIYGTNDIRSEVPEPEIKNPNDVYKIKSYGICGTDLKEFTYSGVPFPPKQGT 60
1 MKARFEDKDIRI-EDIPETVA-PCTVGINVAMCGICGTDLHEF--MGPIFIPCGH 56
61 KDRISGEPLPCGHEFSGTVVEGSGVTSVKPGDRVAE-----ATSHCSDRSRK 112
57 PPHISGSAAVYTMGHEFSGTVYVAVGEGVDIKQAHVYVEPIIHDVPIGESSNHLK 116
113 DVAQDLGLCMACQSGSPNCASLSPFCGLGASGGAFAVYVGEDHVKLPDIPDDIGA 172
117 D-----MNFILGCGGGLSEKIAVRRWHPISDKIPIDQAA 154
173 LVPEISVAMHVERARFQPGTALVLAGGPIG--LATITLQGHAGKIVCSPEALIRQ 230
155 LIPELSTGHAYVRSKAKEDVALVGGAGPIGLLAAVLAKKI---KVITTELSTARKD 211
231 FAKELGAE--VDPSTCDDANAVLKAMPENEGFHAAPDCSGVPQFTTSIVATGPGSIA 288
212 KARESGVADYIILDPSEVDVAEYVK--LTNGEGVDVAEECTSVNKVLDLTVEACKPAANL 269
289 VNAVWGDHPHIGFPMPSLTQOEKATGSMCTYTKDFQEVVKALEDGLISIDKARKMITGK 348
270 VIYSIV-SHPATVNVSHVYVAKELDVAGTIAV-CNDHAETIKLVEEKRINLE---PITQR 324
349 VHLKDGVEKFKOLIEKKNVILYTPN 377
325 IKDELVSKEGFERLIHNNESAVKIIVSPN 353

RESULT 8

H81925
Probable alcohol dehydrogenase (EC 1.1.1.-) MMA0808 [similarity] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81925
R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: H81925
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84091.1; PID:g737952
A:Experimental source: serogroup A, strain 22491

C:Genetics:
A:Gene: MMA0808
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase

Query Match 26.5%; Score 534; DB 2; Length 354;
Best Local Similarity 35.7%; Pred. No. 2.1e-33;
Matches 139; Conservative 55; Mismatches 147; Indels 48; Gaps 12;

1 MGGLIYGTNDIRSEVPEPEIKNPNDVYKIKSYGICGTDLKEFTYSGVPFPPKQGT 60
1 MKARFEDKDIRI-EDIPETVA-PCTVGINVAMCGICGTDLHEF--MGPIFIPCGH 56
61 KDRISGEPLPCGHEFSGTVVEGSGVTSVKPGDRVAE-----ATSHCSDRSRK 112
57 PPHISGSAAVYTMGHEFSGTVYVAVGEGVDIKQAHVYVEPIIHDVPIGESSNHLK 116
113 DVAQDLGLCMACQSGSPNCASLSPFCGLGASGGAFAVYVGEDHVKLPDIPDDIGA 172
117 D-----MNFILGCGGGLSEKIAVRRWHPISDKIPIDQAA 154
173 LVPEISVAMHVERARFQPGTALVLAGGPIG--LATITLQGHAGKIVCSPEALIRQ 230
155 LIPELSTGHAYVRSKAKEDVALVGGAGPIGLLAAVLAKKI---KVITTELSTARKD 211
231 FAKELGAE--VDPSTCDDANAVLKAMPENEGFHAAPDCSGVPQFTTSIVATGPGSIA 288
212 KARESGVADYIILDPSEVDVAEYVK--LTNGEGVDVAEECTSVNKVLDLTVEACKPAANL 269
289 VNAVWGDHPHIGFPMPSLTQOEKATGSMCTYTKDFQEVVKALEDGLISIDKARKMITGK 348
270 VIYSIV-SHPATVNVSHVYVAKELDVAGTIAV-CNDHAETIKLVEEKRINLE---PITQR 324
349 VHLKDGVEKFKOLIEKKNVILYTPN 377
325 IKDELVSKEGFERLIHNNESAVKIIVSPN 353

RESULT 9

B83133
Probable alcohol dehydrogenase (Zn-dependent) PA4097 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83133
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.;
Adman, S.; Yvan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laidig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: B83133
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <STO>
A:Cross-references: GB:AE004826; GB:AE004091; NID:g9950296; PIDN:AA07484.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics: PA4097
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 22.6%; Score 455.5; DB 2; Length 352;
Best Local Similarity 33.5%; Pred. No. 2.2e-27;
Matches 127; Conservative 55; Mismatches 162; Indels 35; Gaps 12;

1 MGGLIYGTNDIRSEVPEPEIKNPNDVYKIKSYGICGTDLKEFTYSGVPFPPKQGT 60
1 MKALRHWARDRLSEL-ERQAPRGVELELVAVACGICGSDLHE--YQSGHSP-QAE 55
61 KDRISGEPLPCGHEFSGTVVEGSGVTSVKPGDRVAEATSHCSDRSRKRYDDYAQDGL 120
56 AHPILSCRAPLTLGHGFCGVAVALPGVGGPRTIGDRAVAEPEYRC-----G 101
121 LCMACQSGSPNCASLSPFCGLGASGGAFAVYVGEDHVKLPDIPDDIGALVPEISVA 180

```

Db 102 ECRYCREGRYNLCESMGFTGLMG-DGFAERARVPATVHLPLPAVGFQRAVALEPAPA 160
Oy 181 WHAVRARFQPGCTALVVGSGPIGLATLALQGHAGKYCSEPALIRROFAKEGA-EV 239
Db 161 LHALKRSLAGQRCVAGVGLIGILLVLMRLRGIEDIAAVDVSPERLALAGEFGSRA 220
Oy 240 FDPSTCDANAVLKAMPENEGFHAAPFCGVPQTFSTIVANGPSIAVNAVWGDHPT 299
Db 221 LDARDGDTA-----ARLRGALDCAFEAAGSASLDAAALASLRKGGELVLSLME--V 273
Oy 300 GFMSMSLTQOKYATGSMCTVVD-FOEYVKALEDGLISDKARKMTTGKVKHLEKDEK 358
Db 274 RLDFPDLVNRRLRLGVSVG--RDAYPELIALALDGRDLARA---VTRSRVLEQAVENG 328
Oy 359 FKQLEIKENNVKILVTPN 377
Db 329 FEALLRDK-SQLKVLVNP 346

```

RESULT 10

```

AH3139
Zinc-binding dehydrogenase Atu4740 [Imported] - Agrobacterium tumefaciens (strain C58, D
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH3139
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L
erge, G.; Gilbert, W.; Grant, C.; Genthner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCelli
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH3139
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <KUR>
A:Cross-references: GB:AE006889; PIDN:AAU45534.1; PID:q17743246; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4740
A:Map position: linear chromosome
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

```

```

Query Match 21.0%; Score 422.5; DB 2; Length 357;
Best Local Similarity 28.5%; Pred. No. 7, 6e-25;
Matches 109; Conservative 75; Mismatches 160; Indels 39; Gaps 10;

```

```

Oy 1 MKGLLYGTNDIRSEVPEPEIKNPNDVKIKSYGICGTDLKEFTYSGPVFPKQGT 60
Db 1 MRLRFHAAKDLRI-EDIAEPKRPQGVLYVNRFPVIGICTDLHEYSY--GPIFIPT 55
Oy 61 KDKISGYELPLCPGHEFSGVTVSVKGGDRVAE-----ATSCSDRSRKYKDT 114
Db 56 PHPFTGAHGPQILGHEFGVVEAIGDGVTSVNVGDRVSIQPLMPRSGDFADRGFL 112
Oy 115 VAODLGLMACQSGSPNCASLFCGLGASGFAEYVYVGGDHMKLPDPSIPDDIGALV 174
Db 113 -----HLSTOLALVGLSWDGGMAEALVNVYOKIDEMTDEALV 156
Oy 175 EPISVAMHAYERARFQGTALVGGPIGLATLALQGHAGKYCSEPALIRROFAKE 234
Db 157 EPSAVAVYACDRGVTAGNSVLTGAGPIGLMTLLAARAAGATQLFVSDLNDARLELARN 216
Oy 235 LGAEV--FDPSTCDANAVLKAMPENEGFHAAPFCGVPQTFSTIVANGPSIAVNAV 292
Db 217 VINDVTTIMPKR-DNVGADVRSATGKVCDAVIECVGNEHMLKACVDVAKQGVVOTG 275
Oy 293 VW-GDHPIGFMPMSLTQOKYATGSMCTVVD-FOEYVKALEDGLISDKARKMTTGKVL 351
Db 276 LHPHNPIDW--FOYTFRDLKIGSMWAPRTHYWPVIRLIASGL---PATKIYTKRITL 330

```

```

Oy 352 KQGEKGFQKLEIKENNVKILV 374
Db 331 DPAVTEGFDALDPAQTHILKILI 353

```

RESULT 11

```

D98148
(R,R)-butanediol dehydrogenase (EC 1.1.1.4) [Imported] - Agrobacterium tumefaciens (
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: D98148
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: D98148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK80710.1; PID:q15158444; GSPDB:GN00170
C:Genetics:
A:Gene: AGR 1.281
A:Map position: linear chromosome
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: oxidoreductase

```

```

Query Match 21.0%; Score 422.5; DB 2; Length 397;
Best Local Similarity 28.5%; Pred. No. 8, 7e-25;
Matches 109; Conservative 75; Mismatches 160; Indels 39; Gaps 10;

```

```

Oy 1 MKGLLYGTNDIRSEVPEPEIKNPNDVKIKSYGICGTDLKEFTYSGPVFPKQGT 60
Db 41 MRLRFHAAKDLRI-EDIAEPKRPQGVLYVNRFPVIGICTDLHEYSY--GPIFIPT 95
Oy 61 KDKISGYELPLCPGHEFSGVTVSVKGGDRVAE-----ATSCSDRSRKYKDT 114
Db 96 PHPFTGAHGPQILGHEFGVVEAIGDGVTSVNVGDRVSIQPLMPRSGDFADRGFL 152
Oy 115 VAODLGLMACQSGSPNCASLFCGLGASGFAEYVYVGGDHMKLPDPSIPDDIGALV 174
Db 153 -----HLSTOLALVGLSWDGGMAEALVNVYOKIDEMTDEALV 196
Oy 175 EPISVAMHAYERARFQGTALVGGPIGLATLALQGHAGKYCSEPALIRROFAKE 234
Db 197 EPSAVAVYACDRGVTAGNSVLTGAGPIGLMTLLAARAAGATQLFVSDLNDARLELARN 256
Oy 235 LGAEV--FDPSTCDANAVLKAMPENEGFHAAPFCGVPQTFSTIVANGPSIAVNAV 292
Db 257 VINDVTTIMPKR-DNVGADVRSATGKVCDAVIECVGNEHMLKACVDVAKQGVVOTG 315
Oy 293 VW-GDHPIGFMPMSLTQOKYATGSMCTVVD-FOEYVKALEDGLISDKARKMTTGKVL 351
Db 316 LHPHNPIDW--FOYTFRDLKIGSMWAPRTHYWPVIRLIASGL---PATKIYTKRITL 370
Oy 352 KQGEKGFQKLEIKENNVKILV 374
Db 371 DPAVTEGFDALDPAQTHILKILI 353

```

RESULT 12

```

A45052
L-Iditol 2-dehydrogenase (EC 1.1.1.14) - Bacillus subtilis
N:Alternate names: glucitol dehydrogenase; polyol dehydrogenase; sorbitol dehydrogen
C:Species: Bacillus subtilis
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 03-Nov-2000
C:Accession: A45052; E69638; I39862
R:Ng, K.; Ye, X.C.; Wong, S.L.
J. Biol. Chem. 267, 24989-24994, 1992
A>Title: Sorbitol dehydrogenase from Bacillus subtilis. Purification, characterizati
A:Reference number: A45052; M0ID:93094198; PMID:1460002
A:Accession: A45052
A:Molecule type: DNA

```

A:Residues: 1-353 <NGI>
A:Cross-references: GB:M69647; NID:g304152; PIDN:AAA22508.1; PID:g304153
A:Experimental source: strain W168
A:Note: Sequence extracted from NCBI backbone (NCBI:120108, NCBI:120111)
A:Note: sequencing of the amino end of the mature protein indicated that Met-1 is removed
A:Note: presents arguments that the third ligand of the catalytic zinc is 156-Glu
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capriano, V.; Carter, N.M.; Chd
A.; Enllich, S.D.; Emerson, P.T.; Enlian, K.D.; Erlington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Hiltz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
A.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
lecl, J.; Hatwood, C.R.; Knecht, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
Koetter, P.; Konungstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69638
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-353 <RUN>
A:Cross-references: GB:A09107; GB:AL009126; NID:g2632866; PIDN:CAB12434.1; PID:g2632928
A:Experimental source: strain W168
R:Ye, R.; Wong, S.L.
J. Bacteriol. 176, 3314-3320, 1994
A:Title: Transcriptional regulation of the *Bacillus subtilis* glucitol dehydrogenase gene
A:Reference number: 139862; MUID:94253000; PMID:8195086
A:Accession: 139862
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-8 <ER>
A:Cross-references: GB:L16626; NID:g436962; PIDN:AAA20875.1; PID:g530793
C:Genetics:
A:Gene: gutB
C:Complex: homotetramer
C:Function: catalyzes the oxidation by NAD+ of sorbitol to fructose
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: homotetramer; NAD; oxidoreductase; zinc
F:30-340/Domain: long-chain alcohol dehydrogenase homology <IADH>
F:175-204/Region: beta-alpha-beta NAD nucleotide-binding fold
F:45-70/Binding site: zinc, catalytic (Cys, His, Glu) #status predicted
F:100,103,106,114/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 20.3%; Score 408.5; DB 1; Length 353;
Best local similarity 27.9%; Pred. No. 8.9e-24;
Matches 109; Conservative 66; Mismatches 155; Indels 61; Gaps 11;

QY 1 MKGLIYGTNDIRYSETPPEIKNDVKIKVSYCGICGNDLKEFTYSGGVFFPKQGT 60
DB 9 MKAAVHNHREIKI-ETLPVPDI-NHDEVLIKVMAGVIGISDILHY-----T 53
QY 61 KDKISGT--ELPLCPGHEPSGTVKPGDRVAEATSHCSRSRKYDTVAOD 118
DB 54 NGRIGNVYVERPPLIGHECAGETAAVAGSSVDQFKVDRVAEAGVTC----- 100
QY 119 LGICMACSSPNCASLSPGIGAGSAGFAEYVYVGGEDHMKLPDIPSTDDIGALVEPS 178
DB 101 -GREACKERKYNLCPVYGLATPRVDGAFVYIKRKODFVFLIPSLSYEEALLEPS 159
QY 179 VAMAVARARFQGTALVYGGPIGATLIALOGHNAKIVSEPALIRPFAKELGAE 238
DB 160 VGIIHAARTKLPQSTIAIKMGVGVIAVAAKAGAGITITDLEPLRLFAKKMGAT 219
QY 239 VFDSPTCDANAVLKKAVPEEGFHAFFDCSGVPQFTTSIVATGDSG----- 286
DB 220 HIINIREDDALEIKI-TINDRGVDVAMETAGNPAALQSLASVRRGKLAIVGLPSOM 278

QY 287 IAVNAVWGDHPHIGEMPSLTYOEKATGSMCYTKDFOEYKALEDLGLSLDRARKMT 346
DB 279 IPLNVFPIIDNEDNEDIYGI-FRIYANTYPPK-----IEFLASIVD---TKHLVT 322
QY 347 GKVHLWDYEGKFKOLIEHKNVILTPN 377
DB 323 DQYSLQ-TQDAMERLQKNECKVMYYPN 352

RESULT 13
T23889
hypothetical protein R04B5.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T23889
R:Wilkinson, J.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19814
A:Accession: T23889
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-347 <WLL>
A:Cross-references: EMBL:Z70782; PIDN:CAA94841.1; GSPDB:GN00023; CESP:R04B5.5
A:Experimental source: clone R04B5
C:Genetics:
A:Gene: CESP:R04B5.5
A:Map position: 5
A:Insertions: 62/3; 138/2; 221/2; 259/3; 300/2
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 19.2%; Score 386.5; DB 2; Length 347;
Best local similarity 29.6%; Pred. No. 4.3e-22;
Matches 112; Conservative 62; Mismatches 151; Indels 53; Gaps 12;

QY 7 YGTNDIRYSETPPEIKNDVKIKVSYCGICGNDLKEFTYSGGVFFPKQGTAKKISG 66
DB 11 YGVDDLR-L-EQVPIRK-PQPNVVLKVTGICGSDVYHNGAIGFVYKE----- 60
QY 67 YELPLCPGHEFSGTVKPGDRVAEATSHCSRSRKYDTVAODLGLCMACQ 126
DB 61 --PMVIGHETGIVSEVNEKHLKVGDRIMEGLPK-----LCECK 103
QY 127 SGSPNCASLSPGIGAGSAGFAEYVYVGGEDHMKLPDIPSTDDIGALVEPSVAMHAVER 186
DB 104 TGRYNLCPEKRFPAIPVAGTISRFPVHDADCFKLPNLSHEDALTEPSVIMHACR 163
QY 187 ARFPQGTALVYGGPIGATLIALOGHNAKIVSEPALIRPFAKELGAEVDPSTCD 246
DB 164 GNVOMGHVYLVLAGPVGIVNLITAKAVAGAVVITDLDGRALAKKLGAD---ATIN 219
QY 247 DANNAVILKAMPE-----NEGFHAFDCSGVPQFTTSIVATGDSGIVANVWGDH--- 297
DB 220 VKGSLDAVASELITLALGDQPPVCEGGAOPSIFATITTKSGVIVLGLADRYEI 279
QY 298 PIGFPMSTLYOEKATGSMCYTKDFOEYKALEDLGLSLDRARKMTGKYLK-DGEV 356
DB 280 PL-----IESATREVDKMGIFRY-VNCYPALELISGKLNLSGLT-----RAHYLLEEQ 329
QY 357 KGFQKOLIEHKNVILY 374
DB 330 EAFRR--TQKADYIKVFI 345

RESULT 14
S32484
L-Iditol 2-dehydrogenase (EC 1.1.1.14) - silkworm
N:Alternate names: sorbitol dehydrogenase
C:Species: *Bombyx mori* (silkworm)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jun-2000
C:Accession: S32484
R:Niml, T.; Yamashita, O.; Yaginuma, T.
Eur. J. Biochem. 213, 1125-1131, 1993
A:Title: A cold-inducible *Bombyx* gene encoding a protein similar to mammalian sorbitol

A:Reference number: S32484; MUID:97087160; PMID:8504807
A:Accession: S32484
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-348 <N11>
A:Cross-references: EMBL:DJ3371; NID:g217259; PIDN:BAA02634.1; PID:g217260
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the oxidation by NAD+ of sorbitol to fructose
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology
C:Keywords: homotetramer; metalloprotein; NAD: oxidoreductase; zinc
F:35-337/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:170-199/Region: beta-alpha-beta NAD nucleotide-binding fold
F:40,65,151/Binding site: zinc, catalytic (Cys, His, Glu) #status predicted
F:95,98,101,109/Binding site: zinc, noncatalytic (Cys) #status predicted

Query Match 18.4%; Score 370.5; DB 1; Length 348;
Best Local Similarity 30.7%; Pred. No. 7.2e-21;
Matches 103; Conservative 56; Mismatches 146; Indels 31; Gaps 10;

OY 7 YGTNDIRYSETPPEIKNPNDVKIKVSYGICGTDLKEFTYSGPVFFPKQTKDKISG 66
DB 10 HGADVRI-EKIPVPEI-NDEVLIKIDCVIGCSGVK--LYSTGTC-----GADVI-- 57
OY 67 YELPLCGHERSGTVEYSGVTSVKPGDRVAVEATSHCSDRSRKDTVAODLGLCMACQ 126
DB 58 -DKPIYIGHGACGTIVVGVGVSSLRVGRVAILPTQPC--RS-----CELCK 102
OY 127 SGSPNCASLSFCGLGASGGAFAEYVYVYGGEDHMYKLPDIPDITGALVEPTISVAMHAYR 186
DB 103 RGYNLCEPEPRYSSMGAPGMLCRYKHVADFCHKLPDNLMEGAANVQPLAIVIHACNR 162
OY 187 ARFPGGTALVGGCPITGLATITLALOGHHACKIVCSEPALIRQFAKELGAE--VDPST 244
DB 163 AKITLGSKIYVLGAGPIGLICMSAKAMGASKITLIDVOSRLDALELGADNVILVRR 222
OY 245 CDDANAVLKAVPENEGFHAFFDCSGVPQFTTSIVATGPGIAVNVAVMGDHPGEMPM 304
DB 223 YTDDEVVEKIYKLLGDRDVSIDACGYSAQRVALVTKTAGLVVGI-ADKTVE-LPL 280
OY 305 SLTQEKYATGSMCTYKDFQEVVKALEDGLISDK 340
DB 281 SQALLREVDDVVSFRIMNTYQPALAVSSGAIPDK 316

RESULT 15
B72381
alcohol dehydrogenase, zinc-containing - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: B72381
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gilm, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garetty, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.J. C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72381
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-395 <ARN>
A:Cross-references: GB:AE001720; GB:AE000512; NID:94980906; PIDN:AND3497.1; PID:9498091
C:Genetics:
A:Experimental source: strain MSB8
A:Gene: TM0412
C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 18.4%; Score 370; DB 2; Length 395;
Best Local Similarity 34.0%; Pred. No. 9.2e-21;
Matches 109; Conservative 37; Mismatches 111; Indels 64; Gaps 13;

OY 4 LLYGTNDIRYSETPPEIKNPNDVKIKVSYGICGTDLKEFTYSGPVFFPKQ 58
DB 103 RGYNLCEPEPRYSSMGAPGMLCRYKHVADFCHKLPDNLMEGAANVQPLAIVIHACNR 162

DB 27 LTVLGSKVMRYPEVREVEVEPEPRIEKTETIIIRVAKAGICGSDVHMAQTDDEGYIYP-- 84
OY 59 GTDKDXTSGYELPLCPGHEFSGTVEYSGVTSVKP-----GDRVAVEATSHCSDRSRK 112
DB 85 ----GLTGF--PYTLGHERSGVYVEAGPEALINRTRKRFETIGEPVCAEMLMC----- 131
OY 113 DTVADLGLCMACOSGSPNCASLSFCGLGASGGAFAEYV-----YGED 157
DB 132 -----GHCRCPEGEPNHCENLNLGF--NVDCAFAYEYKVDKAYAMSLRELEGVEGD 183
OY 158 HMYKLPDIPDITGALVEPTISVAMHA--VERARFPGQTALVIGCGPIGLATITLALOGHH 215
DB 184 RLFL-----AGSLVEPTISVAMNAVIVRGGIRPDNVNVLISGPIGLAVALILKHAQ 235
OY 216 AGKIVCSEPALIRQFAKELGAE--VDPSTCDDANAVLKAVPENEGFHAFFDCSGVPQT 274
DB 236 ASKYILSEPESEVRNKLAKELGADHVIDPTKENFEAVLD--YTNGLGAKLFLEATGVPL 293
OY 275 FTTSI--VATGPGIAVNAV 293
DB 294 VMPQIEEVIWRARGINATVAI 314

Search completed: May 2, 2003, 13:01:43
Job time: 623 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:08:47 ; Search time 239 Seconds

(without alignments)
46,781 Million cell updates/sec

Title: US-10-020-674-2

Sequence: 1 MKGLLYGNDIRSEVTYPE.....OLIEHKNVKITLVPEVYS 380

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_MA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	491.5	24.4	354	4	US-09-134-001C-3374
2	361	17.9	371	4	US-09-347-803-27
3	360.5	17.9	356	3	US-08-466-548B-2
4	360.5	17.9	356	3	US-07-998-326F-2
5	360.5	17.9	356	5	PCT-US93-13560-2
6	349	17.3	316	4	US-09-347-803-20
7	327	16.2	301	4	US-09-347-803-22
8	326	16.2	341	4	US-09-134-001C-16
9	265.5	13.2	382	4	US-09-134-001C-4483
10	265.5	13.2	382	4	US-09-134-001C-3003
11	244.5	12.1	336	1	US-08-713-254-2
12	244.5	12.1	336	4	US-08-855-767-2
13	236	11.7	352	4	US-09-504-358-18
14	236	11.7	352	4	US-09-954-314-18
15	233.5	11.6	374	4	US-09-347-878-52
16	229.5	11.4	374	2	US-07-857-224B-82
17	228.5	11.4	374	2	US-07-857-224B-84
18	225.5	11.2	376	1	US-08-608-241-2
19	225.5	11.2	376	2	US-08-922-182-2
20	225.5	11.2	376	2	US-08-919-953-2
21	225.5	11.2	376	4	US-09-192-983-2
22	225.5	11.2	386	4	US-09-370-838-81
23	223.5	11.1	374	2	US-07-857-224B-83
24	222.5	11.1	375	4	US-09-347-878-56
25	222	11.0	346	2	US-07-857-224B-96
26	222	11.0	375	2	US-07-857-224B-86
27	221.5	11.0	377	2	US-07-857-224B-90

28	218	10.8	344	4	US-09-134-001C-5343	Sequence 5343, Ap
29	214	10.6	374	2	US-07-857-224B-85	Sequence 85, Appl
30	213.5	10.6	376	2	US-07-857-224B-91	Sequence 91, Appl
31	208.5	10.4	349	2	US-07-857-224B-95	Sequence 95, Appl
32	206	10.2	374	4	US-09-347-878-54	Sequence 54, Appl
33	204	10.1	379	2	US-07-857-224B-87	Sequence 87, Appl
34	201.5	10.0	374	2	US-07-857-224B-80	Sequence 80, Appl
35	201.5	10.0	374	2	US-07-857-224B-81	Sequence 81, Appl
36	195.5	9.7	376	2	US-07-857-224B-89	Sequence 89, Appl
37	194.5	9.7	352	2	US-08-805-191-2	Sequence 2, Appl1
38	190.5	9.5	376	2	US-07-857-224B-88	Sequence 88, Appl
39	189.5	9.4	343	4	US-09-363-189B-4	Sequence 4, Appl1
40	189	9.4	379	3	US-08-937-610-2	Sequence 2, Appl1
41	185.5	9.2	344	2	US-07-857-224B-93	Sequence 93, Appl
42	179.5	8.9	345	2	US-07-857-224B-94	Sequence 94, Appl
43	170.5	8.5	344	2	US-07-857-224B-92	Sequence 92, Appl
44	148	7.4	93	4	US-08-936-165A-46	Sequence 46, Appl
45	147.5	7.3	396	4	US-08-860-656B-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-134-001C-3374
; Sequence 3374, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3374
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3374

Query Match 24.4% Score 491.5; DB 4; Length 354;

Best Local Similarity 34.9% Pred. No. 3.5e-43;
Matches 132; Conservative 59; Mismatches 154; Indels 33; Gaps 11;

QY	1	MKGLLYGNDIRSEVTYPEPEIKNPDKIKVSGICGTDLKEFTYSGGPFPPOGT	60
DB	5	MKAAMVIGQKDVREDR-EPKAIKDNEVOVKSMAGICGTDLHE--YLESPIT-STDQ	59
QY	61	KDKISGYELPLCPGHEFSGVTEVSGVTSVPCDRAVAENTSHCSDRSRYKDTVAODLG	120
DB	60	PDPLIGQTAPVTLTLEHSGVVENGVKNVSRKKDRVAVNT--VSKREKENVLDV--	115
QY	121	LCMACGSGSPNCASLSFCGICGASGFAEYVYVGEDHMKVLPISIPDDIALVEPISVA	180
DB	116	-----GYSFTIG-GSDGAFETNPAPETNVYLPNVNSAREBALVEPAVA	160
QY	181	WHAVERARFOPGOTATLVGGGPIGLATILAGHHAGKIVSEPALTRORAKELGA-EV	239
DB	161	VOAVKEGELFLGDTVAVFGACPIGLTITVAKAAGASKIFVDSERLAKSVGATHV	220
QY	240	FDPSTCDANAVLKAAMPENGFHAAPDCSGVPOFTTSYATGPGSIAVNAVAGCHI	299
DB	221	YNSGNVDPVQTVYE--HTDNGVDVSFEVAGVITLQOSIEVTRPRGTAIVSIFG-HPV	276
QY	300	GFMPSMLTQKRVATGSKVTKPOEVYKALEGLISLDAKAKMITGKVLKQGVENG	359
DB	277	EFNPLQMNKGVKLTITVAVPTTFQOTIDLIANGSLNV--KDVYTDQIDLDNIVESG	333


```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/998,226F
FILING DATE: 30-DEC-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WALSH, ANDREA C
REGISTRATION NUMBER: 34988
REFERENCE/DOCKET NUMBER: 5283.us.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-998-226F-2
```

```
Query Match          17.9%; Score 360.5; DB 3; Length 356;
Best Local Similarity 30.3%; Pred. No. 2,1e-29;
Matches 11; Conservative 56; Mismatches 160; Indels 39; Gaps 11;

QY 16 EYPEPEIKNPNDYKIKVSYCGICGTDLKEFTYSGGVPFPKQGTDKISGYEL--PLCP 73
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 20 EQRPIPTIKDPHYKLAIKATGICGSDI-HYRSGG-----IGKYLKAPMVL 66

QY 74 GHEFSGVIVVSGSVTSVKRGDRAVAEATSHCSDRSRYKKTVAADLGLCACSGSPNCC 133
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 67 GHSSGCGVVEGDAVTVKVGDRVALEP---GVPSRYSDET-----KEGRYNLC 112

QY 134 ASLFCGLGASGSGFAEYVYVGEDHMYKLPDSIPDDIGALVEPISVAMHAEARFPQ 193
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 113 PHMAFATPTPIDGLTKVYKYSPEDFLVKLPBGVSYEGACVEPLSGVSHNKLAVRFGT 172

QY 194 TALVGGPPIGLATILALQGHNAKIVCSEPALIRQFAKELGA-EVFDP--TCDDANA 250
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 173 KVVVFGAGPYGLLTGAVARAFGATDVIFVDVFNKLGQRADFCATNTFNSSQSTDKAOD 232

QY 251 VLKAMPENEGFHA--AFDCSGVPQTFSTTSIVATGPGSIIVNVAWGDHPIGFMPSLTY 308
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 233 LADGVOKLLGSHNADVFECSGADVCIDAAVKTKVGGTMOVGW-GKNTNPIAEVSG 291

QY 309 QEKYATGSMCYTKDFQEVYKALEDLGLISDKARKMITGKVHLKDGVEKGFQOLIEHEN 368
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 292 KEMKLIGCFRYSFGDYRDVAVNLVATGKVNV---KPLITHKFKEDAA-KAYDYNIAHGE 347

QY 369 NVKILV 374
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 348 VVKTTI 353
```

```
RESULT 5
PCT-US93-12560-2
Sequence 2, Application PC/TUS9312560
GENERAL INFORMATION:
APPLICANT: sartby, aarna v
APPLICANT: schopp, cythla w
TITLE OF INVENTION: ENHANCED YEAST EXPRESSION IN USING
TITLE OF INVENTION: REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESS: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/12560
FILING DATE: 22-DEC-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/998,226
FILING DATE: 30-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: WALSH, ANDREA C
REGISTRATION NUMBER: 34988
REFERENCE/DOCKET NUMBER: 5283.us.01
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 356 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-12560-2
```

```
Query Match          17.9%; Score 360.5; DB 5; Length 356;
Best Local Similarity 30.3%; Pred. No. 2,1e-29;
Matches 11; Conservative 56; Mismatches 160; Indels 39; Gaps 11;

QY 16 EYPEPEIKNPNDYKIKVSYCGICGTDLKEFTYSGGVPFPKQGTDKISGYEL--PLCP 73
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 20 EQRPIPTIKDPHYKLAIKATGICGSDI-HYRSGG-----IGKYLKAPMVL 66

QY 74 GHEFSGVIVVSGSVTSVKRGDRAVAEATSHCSDRSRYKKTVAADLGLCACSGSPNCC 133
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 67 GHSSGCGVVEGDAVTVKVGDRVALEP---GVPSRYSDET-----KEGRYNLC 112

QY 134 ASLFCGLGASGSGFAEYVYVGEDHMYKLPDSIPDDIGALVEPISVAMHAEARFPQ 193
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 113 PHMAFATPTPIDGLTKVYKYSPEDFLVKLPBGVSYEGACVEPLSGVSHNKLAVRFGT 172

QY 194 TALVGGPPIGLATILALQGHNAKIVCSEPALIRQFAKELGA-EVFDP--TCDDANA 250
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 173 KVVVFGAGPYGLLTGAVARAFGATDVIFVDVFNKLGQRADFCATNTFNSSQSTDKAOD 232

QY 251 VLKAMPENEGFHA--AFDCSGVPQTFSTTSIVATGPGSIIVNVAWGDHPIGFMPSLTY 308
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 233 LADGVOKLLGSHNADVFECSGADVCIDAAVKTKVGGTMOVGW-GKNTNPIAEVSG 291

QY 309 QEKYATGSMCYTKDFQEVYKALEDLGLISDKARKMITGKVHLKDGVEKGFQOLIEHEN 368
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 292 KEMKLIGCFRYSFGDYRDVAVNLVATGKVNV---KPLITHKFKEDAA-KAYDYNIAHGE 347

QY 369 NVKILV 374
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 348 VVKTTI 353
```

```
RESULT 6
US-09-347-803-20
Sequence 20, Application US/09347803
Patent No. 6274379
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.
APPLICANT: Hiltz, Bill
APPLICANT: Kinney, Tony
APPLICANT: Orozco, Buddy
TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
FILE REFERENCE: BB-1176
CURRENT APPLICATION NUMBER: US/09/347,803
CURRENT FILING DATE: 1999-07-02
EARLIER APPLICATION NUMBER: 60/092,952
EARLIER FILING DATE: July 15, 1998
NUMBER OF SEQ ID NOS: 27
SOFTWARE: Microsoft Office 97
SEQ ID NO 20
```

```

; LENGTH: 316
; TYPE: PRT
; ORGANISM: Glycine max
US-09-347-803-20

```

Query Match	17.3%	Score 349;	DB 4;	length 316;
Best Local Similarity	33.3%	Pred. No. 2.8e-28;		
Matches 88;	Conservative 36;	Mismatches 112;	Indels 28;	Gaps 4;

```

QY      26  PNDVAKIVSYSGICIGCTDLKEFTTSSGVPVEFPKQTKDKRISGELYELPLCPGHESGTVVEVG
          |||:::  |||:::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      41  PHDVAFVRMKAVGICGSDVHYLKLTRCAHFVKE-----DMVIGHECAGIIIEVG

```

```
QY      86 SCYTSVKKPGDRVAVEATSHCSDSRSRYKDYAQDLGLMACQSSSPNCASLSFGGLGGAS 145
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB     90 SQYKSLVFGDRVALIEPGISCMH-----CNHCKHGRYNLCDMDMKFEATPPVH 135
```

Dy 146 GGTAEEVYVGEDHMYKLPSDISPPDDIGALVEPISVAWMAVERARQPGQTALVLGGSPIGL 205
| : | : |||| : || : || | | | : || ||||
Db 136 GSLANQIHPADLCERKLPDNYLSLEEGAMCEPLSGVGHACRANGPETNVLIMGAPIGL 195

QY 206 ATTALOGHHAGKIVCSEPALIRROFAKELGA-EVFDEST--CDANAVLKAMPENEGF 262
| : | : | : | : | : | : | : | : | : |
DB 196 VTMLAARAFGAPRTYIVDVDDHRLSVAKSLGADDIKVSTNIKVAEEVVOIQKWGAGI 255

QY	263 HAAFDSCGVPQTFTTSIVATGPSG	286
	: : : ::	
Db	256 DVTFCAGFGDKTMSPALSATQPGG	279

RESULT 7
US-09-347-803-22

Patent NO. 6274379
GENERAL INFORMATION:
APPLICANT: Farnodu, Layo O.

APPLICANT: Kinney, Tony
 APPLICANT: Orozco, Buddy
 TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes

```

; CURRENT APPLICATION NUMBER: US/09/347,
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952

```

```

;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: Microsoft Office 97
;  SEQ ID NO 22

```

```

; LENGTH: 301
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-347-803-22

```

Query Match	16.28;	Score 327;	DB 4;	Length 301;
Best Local Similarity	31.5%;	Pred. No. 5.4e-26;		
Matches 84; Conservative	42;	Mismatches 107;	Indels 34;	Gaps 5;

QY 26 PNDVRIKVSICIGTD---LKEFTYSGPVFFPKQGTDKDISGELPLCPGHEFSGTVV 82

Db 25 PYDVRIRMKAVSICGSDHYILKERRIAH---FVVKR-----PWVIGHEAGIIE 70

QY 83 EVGSGVTSVKPGRVAVEATSHCSDRSKRYDITVAQDLGLMACQSSPNCASLSFCGLG 14222

QY 143 GASGFAEYVYGEDHMKLPDISPDDIGALVEDISVANHAVERARFQPCCATLVLCGGP 2022

D5 117 RYNGSIAADOTVVYBDCI.CEKI.DNNVSI.FEGCAMCEPT.SVGVIACBRADVGA.EKSVU.IIMAGPD 176

203 IGLATIALOGHHAGKIVCSEBALIRQFAKETGAEV--FDPSTCDDANAVLKAWPEN 259
 177 TCTYMTGSAPECAPIVTDNDNDHPTGVAKCTGADAVIVCSGNEPDIAGCTEPTOAMG 238
 178

```
QY      260 EGFHAIFDCSGVPQTFTTSIVATGPGS 286
      :  ||:  : ||:  ||:  ||:  ||:  ||:
Db      237 GDIDVSLDCAGFSKTMSTALEATRPGG 263
```

RESULT 8
US-09-347-803-16

Patent NO. 6274379
GENERAL INFORMATION:
APPLICANT: Famodu, Layo O.

APPLICANT: Kinney, Tony
APPLICANT: Orozco, Buddy
TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes

```

; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952

```

```

;      NUMBER OF SEQ ID NOS: 27
;      SOFTWARE: Microsoft Office 97
;      SEQ ID NO 16

```

US-09-347-803-16

Query Match	16.2%	Score 326;	DB 4;	Length 341;
Best Local Similarity	29.7%	Pred. No. 8.4e-26;		
Matches 103; Conservative	47;	Mismatches 137;	Indels 60;	Gaps 10;

```

QY      26 PNDVKIKVSYCGICGTD---LKEFTYSGGPFPPKGGTKDKISGYELDLCPGHEFSGTV 82
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      33 PYDVRVRMKAAGVIGCGSDVHYLRRIARH---FVWKE-----DMVIGHCAGVVE 78

```

QY 83 EVGSGVTSVKPGDRAVAEATSHCSDRSRKYKDTVAQDLGLCMACQGSGSPCCASISFCGLG 14
|||::| : ||||: | : | : | :
Db 79 EVGAGTHTLSVGDRVALKEPVS-----WRCRHCKGGGRYPVRNMKEFATP 12

QY 143 GASGFAEYVXXGEDHMYKLPDSDIPDDIGALVERISYAMHAVERAREREGOTALVLGGP 20
- - - - -
Db 125 PVHGSLANQVNIHPADLCFKLPDGSLEEGACNCEPLSGVNHACRAGAVPETGLVVGAGR 18

```

QY      203 IGLATILALQGHAGKIVCSEPALIRQFAKELGAEV--FDPESTDANAVLKAMPEN 25
      |||::||:  -:::  -:::  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db      185 IGLVSLIARACGAPRVVVVDVDDHRLAVARSLGADAAVRSRAEDLADEVERIRAMG 24

```

DQ 260 EGFHAFDCCGVPQTFTTSIVATGSGIAVNVAVMGDPHPIGENPMSLTREQKATGSMCY 31
:
: : | : : : : : : : : : : :
Db 245 SDIDVSIDCAGSFKTSMSTALEATRGGCKVLVGMGHE---MTLPPLAREVDVVGVFRY 30

```
QY      320  TVKDFQEVKALEDLGLSLDKARKMITGKVHLLKDGVEKGFQOLIEHK 366
      11      : : 1 1 : : : : 1 1 1 :
Db      301  --KDTWP-----LCIDFLR--SGKVDV-----KPLTHRH 325
```

RESULT 9
US-09-134-001C-4483

; Sequence 4483, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Ivern Doucette-Stamm et al]

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO SIMPHILLOCOCCUS
 TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: ITS/09/134-001C

```

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,776

```

; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674

; NUMBER OF SEQ ID NOS: 56/4

SEO ID NO 4483
 LENGTH: 382
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-4483

Query Match 13.2% Score 266.5; DB 4; Length 382;
 Best Local Similarity 25.5%; Pred. No. 1.8e-19;
 Matches 97; Conservative 63; Mismatches 145; Indels 75; Gaps 16;

QY 20 EPEIKNPDKIKVSYGICGTDLKEFTYSGSPVFPKQGTCKKISGVELPLCPGHEFSG 79
 DB 51 KPIITSTDAIIRVKTITIGTDL-----HIKGTPEVKSHTL---GHEGIG 96
 QY 80 TVEVSGSVSVKPGDRVAEATSHCSDRSRKYDTVAODIGLCMACOSSGSPNCASISFC 139
 DB 97 IIEIGDNVNNFKVGDVVIISCISSC-----GKCYCKKG-----IYAHC 136
 QY 140 GLG-----ASGGEFNEY--VYGEDHMYKLPDIPDDIGALVEPISVAMH--AVERAR 188
 DB 137 ENGGMILGHLVNGTQAEVKKVPADNSLYHAPSNLKDDALVMSDILPLTGYEIGVLK 196
 QY 189 FQPGQALVLGGFPIGATIALQGHAGKIVGSEPALIRROFAKELGA-EVDPSTCDD 247
 DB 197 VKGCTVAIYGAGPVGLAALLTAQFYSKIIMTDLDNLETKELGATHLINSKETET 256
 QY 248 ANAVLKAMPENEGEHAAPDCGVPOTFTTSIVATGPGSIAVNAVWGDHPICGMPSLT 307
 DB 257 AIKKVKSLNP--RCVDVAIEAVGIPOTFDLQNLIGVDGDIANVAGH-----LPVOLD 308
 QY 308 YQEKY-----ATGSMCYTKADEQEVYKALDGLISLQKARKMTGTVHKLQDVEKG-- 358
 DB 309 IDKIMININVTGLV--SGNTHEELLEALKSKIT---QPEQLVHYSKLSF-IESAYDL 362
 QY 359 FKOLIEHK-----ENNVKI 372
 DB 363 FRNATDHKAIRKLIENDITI 382

RESULT 10
 US-09-134-001C-3003
 ; Sequence 3003, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3003
 ; LENGTH: 381
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-3003

Query Match 13.2% Score 265.5; DB 4; Length 381;
 Best Local Similarity 24.4%; Pred. No. 2.3e-19;
 Matches 96; Conservative 54; Mismatches 152; Indels 91; Gaps 13;
 QY 14 YSETVPEP---EIKNP--NDVKIKVSYGICGTDLKEFTYSGSPVFPKQGTCKKISGY 67
 DB 23 YKESKPLKIEYLELDNPEHEHLKIRHAAGCHSDLSVI-----NGNRP----- 67
 QY 68 ELPLCPGHEFSGVYGVSGTSVKKPGDRVAEATSHCSDRSRKYDTVAODIGLCMACOS 127
 DB 68 PLPALGHEASGVKAVTRVSEGDHVVCTPIPSC-----GKCIPEKE 113

QY 128 GSPNCASISFCGLG-----ASGGEFNEYVYGEDHMYKLPDIP 167
 DB 114 GRALCENGAISNEKGEMLGEMRLSDECKYVHHLGISGFAEYSVSENSIYKIDKIP 173
 QY 168 DD-----IGALVEPISVAMHVERARFQPGQALVLGGFPIGATIALQGH 215
 DB 174 FERAAFGCAIITIGIAGV-----VNTAQIRSGSNVAVVGLGIGINALIGAKLAG 223
 QY 216 AGKIVSEPLIRROFAKELGAEVFPSTCDDANAVLKAMPENEGEHAAPDCGVPOTF 275
 DB 224 ANELIADINEDFEELAKOGFATFNSSDKDIDQIKKEYIP--GGEVAFETAGVPPAM 281
 QY 276 TTSIVATGPGSIAVNAVWGDHP---IGFPMPSLTQEKYATGS---MCTYKDFQEVK 329
 DB 282 KVAQITKRGGTYTTL--PNKDNFSFQVYTLAEERIKSGSVSCVPDRDIPRFVN 339
 QY 330 ALEDGLISLQKARKMTGTVHKLQDVEKGKQL 362
 DB 340 LYNGGRINIDS---LISEVITLDEINEGFDRL 368

RESULT 11
 US-08-713-254-2
 ; Sequence 2, Application US/08713254
 ; Patent No. 5763236
 ; GENERAL INFORMATION:
 ; APPLICANT: KOTIMA, TOMOKO
 ; APPLICANT: YAMAMOTO, HIROAKI
 ; APPLICANT: KANADA, NAOKI
 ; APPLICANT: MATSUYAMA, AKINOBU
 ; TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID
 ; TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
 ; TITLE OF INVENTION: CONTRAINING SAID DNA SEGMENT AND A METHOD OF PREPARING
 ; TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ; ADDRESSEE: P.C.
 ; STREET: 1755 S. Jefferson Davis Highway, Suite 400
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/713,254
 ; FILING DATE: 12-SEP-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/311,328
 ; FILING DATE: 23-SEP-1994
 ; APPLICATION NUMBER: JP 5-261649
 ; FILING DATE: 24-SEP-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 5-337191
 ; FILING DATE: 28-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 6-181308
 ; FILING DATE: 02-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Oblon, No. 5763236man F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 6423-001-0
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 413-3000
 ; TELEFAX: (703) 413-2220
 ; TELEX: 248855 OPAT UR
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-254-2

Query Match 12.1%; Score 244.5; DB 1; Length 336;
Best Local Similarity 25.8%; Pred. No. 3e-17;
Matches 92; Conservative 53; Mismatches 146; Indels 65; Gaps 16;

QY 31 IKVSYGICGTDLKEFTYSGGVFFPKQGTDKRISGELPLCGHESGTVVEVSGSVTS 90
DB 36 LKVDVAVGLCHSDL-HVITYEG-----LDCGDNYVM---GHEIAGTVAAVGDDVIN 80
QY 91 VKPDRAVAEATSHCSDRSRYKDTVAADLGLCMACGSGSPNCASL--SFCGLGASGGE 148
DB 81 YKVGDRVACVPGNGC-----GGCKYCCGALDNCNKAQSPNLLIGAGGLGN 125
QY 149 AEY-VVYGEDHMKLPDSIPDDIGAL-VEPISVAHVAERARPOQTALVYGCGPIGLA 206
DB 126 QQYLIVTRPRNLRIPIDYNADVAASDAVLPYHAIKMAQVSPNLLIGAGGLGN 185
QY 207 TIALGHNAGKIVCSEPALIRROFAKELCAEYFDPSTCDANAVLKAMPEN---EGFH 263
DB 186 AIQVAKAFGA-KVTVLDKKRKARDQAKKLGD-----AYETL-PESISPGSFS 232
QY 264 AAFDCSGVPQFTTTSIVATGPGSIAVNAVWGDHPIGFPMPSLTYOEKATGSMCTYVD 323
DB 233 ACDFEVSQATFVOCQKYVEPKVIMPVGL-GAPNLSFNLGDLALREIRILISFGWGTND 291
QY 324 FOEVKALDEGLISLDKARKMITGKVHLKDGVEKGFOLIEHKNV---KILVTP 376
DB 292 LDDVLKLVSEGV-----KPVYRSKAKE-----LPEYTEKLRNNAYGRVAVFP 336

RESULT 12

US-08-853-767-2
Sequence 2, Application US/08855767
Patent No. 6255092

GENERAL INFORMATION:

APPLICANT: KOJIMA, TOMOKO
APPLICANT: YAMAMOTO, HIROAKI
APPLICANT: KAWADA, NAOKI
APPLICANT: MATSUYAMA, AKINOBU
TITLE OF INVENTION: NOVEL ENZYME, A METHOD TO PREPARE SAID
TITLE OF INVENTION: ENZYME, A DNA SEGMENT ENCODING SAID ENZYME, A TRANSFORMANT
TITLE OF INVENTION: COMBINING SAID DNA SEGMENT AND A METHOD OF PREPARING
TITLE OF INVENTION: OPTICALLY ACTIVE ALCOHOL USING SAID ENZYME
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/855,767
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,328
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-337191
FILING DATE: 28-DEC-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 6-181308
FILING DATE: 02-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, No. 6255092man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 6423-001-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 336 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-853-767-2

Query Match 12.1%; Score 244.5; DB 4; Length 336;
Best Local Similarity 25.8%; Pred. No. 3e-17;
Matches 92; Conservative 53; Mismatches 146; Indels 65; Gaps 16;

QY 31 IKVSYGICGTDLKEFTYSGGVFFPKQGTDKRISGELPLCGHESGTVVEVSGSVTS 90
DB 36 LKVDVAVGLCHSDL-HVITYEG-----LDCGDNYVM---GHEIAGTVAAVGDDVIN 80
QY 91 VKPDRAVAEATSHCSDRSRYKDTVAADLGLCMACGSGSPNCASL--SFCGLGASGGE 148
DB 81 YKVGDRVACVPGNGC-----GGCKYCCGALDNCNKAQSPNLLIGAGGLGN 125
QY 149 AEY-VVYGEDHMKLPDSIPDDIGAL-VEPISVAHVAERARPOQTALVYGCGPIGLA 206
DB 126 QQYLIVTRPRNLRIPIDYNADVAASDAVLPYHAIKMAQVSPNLLIGAGGLGN 185
QY 207 TIALGHNAGKIVCSEPALIRROFAKELCAEYFDPSTCDANAVLKAMPEN---EGFH 263
DB 186 AIQVAKAFGA-KVTVLDKKRKARDQAKKLGD-----AYETL-PESISPGSFS 232
QY 264 AAFDCSGVPQFTTTSIVATGPGSIAVNAVWGDHPIGFPMPSLTYOEKATGSMCTYVD 323
DB 233 ACDFEVSQATFVOCQKYVEPKVIMPVGL-GAPNLSFNLGDLALREIRILISFGWGTND 291
QY 324 FOEVKALDEGLISLDKARKMITGKVHLKDGVEKGFOLIEHKNV---KILVTP 376
DB 292 LDDVLKLVSEGV-----KPVYRSKAKE-----LPEYTEKLRNNAYGRVAVFP 336

RESULT 13

US-09-504-358-18
Sequence 18, Application US/09504358
Patent No. 6365376

GENERAL INFORMATION:

APPLICANT: Rouviere, Pierre E.
APPLICANT: Brzostowicz, Patricia C.
TITLE OF INVENTION: GENES AND ENZYMES FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIA
FILE REFERENCE: BC1001 US NA
CURRENT APPLICATION NUMBER: US/09/504,358
CURRENT FILING DATE: 2000-02-15
EARLIER APPLICATION NUMBER: 60/120,702
EARLIER FILING DATE: 1999-February-19
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Microsoft Office 97
SEQ ID NO 18
LENGTH: 352
TYPE: PRT
ORGANISM: Brevibacterium sp HCU
US-09-504-358-18

Query Match 11.7%; Score 236; DB 4; Length 352;
Best Local Similarity 28.1%; Pred. No. 2.6e-16;
Matches 104; Conservative 40; Mismatches 154; Indels 72; Gaps 16;
QY 20 EPEIKNP-----NDVAKIVSYGICGTD--LKEFTYSGGVFFPKQGTND-KISGELPLC 72

```

Db 15 EIELDPRKMGREVLKVTNAGVCHTDTHQDDGIDLG-----SRGLDMSTGYVPCV 69
Qy 73 PGHEFSGTVEVCGSVTSKPGDRVAVENTSHCSDRSRYKDTVAODLGICMACQSGSPNC 132
Db 70 MGHETVEVEVEGEDVTDVAAGDTCLAFWIGC-----GECGCAHGHENA 115
Qy 133 CASLFCGAGGAGFAEVVYGEDHMKLPDSIPDDIGAL-VEPISVAMHAV-----184
Db 116 CDNGRALGI-IOFGFAEYIL-----LPDORAIIDVAGVDPMAAATLACSGVT 162
Qy 185 -----ERARFQPGOTALVILGGPRLGATLILALOGHNAGKIYCSSEPALIRROFAKELGA 237
Db 163 SYSSARKATATVPDEPIGVMGVGMVTAVALGHNIIAIDVSDENLASAQLGA 222
Qy 238 EYFDPSTCDANAVLKAMPENEG-FHAAFDCSGVPQFTTSIVATPGSIAVNAVWGD 296
Db 223 TL-----TVSKNATSHDLVLAAGGFIAIIDLVTGDTVALADALSRAKGIYQVGLFEG 278
Qy 297 H---PIGFPM-SLTYOEKATGSMCTVAKDFEYVKALEDGLISLDKARKMITGKYLK 352
Db 279 EFVVPATAMALKGLTLQGNVVG-----TYEVEVEVELARQG--SLPKL--PITGGLTV 329
Qy 353 DGEVKGFKOL 362
Db 330 DGVNDGLERL 339

```

```

RESULT 14
US-09-954-314-18
; Sequence 18, Application US/09954314
; Patent No. 6465224
; GENERAL INFORMATION:
; APPLICANT: Rouviere, Pierre E.
; APPLICANT: Brzostowicz, Patricia C.
; TITLE OF INVENTION: GENES AND ENZYMS FOR THE PRODUCTION OF ADIPIC ACID INTERMEDIATES
; FILE REFERENCE: BC1001 US NA
; CURRENT APPLICATION NUMBER: US/09/954,314
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: 60/120,702
; PRIOR FILING DATE: 1999-February-19
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 352
; TYPE: PRF
; ORGANISM: Brevibacterium sp HCU
US-09-954-314-18

```

```

Query Match 11.7%; Score 236; DB 4; Length 352;
Best Local Similarity 28.1%; Pred. No. 2,6e-16;
Matches 104; Conservative 40; Mismatches 154; Indels 72; Gaps 16;
Qy 20 EBEIKRP-----NDYKIKVSYGICGTD--LKEFTYSGGVFPFKQGTKD-KISGYELPLC 72
Db 15 EIELDPRKMGREVLKVTNAGVCHTDTHQDDGIDLG-----SRGLDMSTGYVPCV 69
Qy 73 PGHEFSGTVEVCGSVTSKPGDRVAVENTSHCSDRSRYKDTVAODLGICMACQSGSPNC 132
Db 70 MGHETVEVEVEGEDVTDVAAGDTCLAFWIGC-----GECGCAHGHENA 115
Qy 133 CASLFCGAGGAGFAEVVYGEDHMKLPDSIPDDIGAL-VEPISVAMHAV-----184
Db 116 CDNGRALGI-IOFGFAEYIL-----LPDORAIIDVAGVDPMAAATLACSGVT 162
Qy 185 -----ERARFQPGOTALVILGGPRLGATLILALOGHNAGKIYCSSEPALIRROFAKELGA 237
Db 163 SYSSARKATATVPDEPIGVMGVGMVTAVALGHNIIAIDVSDENLASAQLGA 222
Qy 238 EYFDPSTCDANAVLKAMPENEG-FHAAFDCSGVPQFTTSIVATPGSIAVNAVWGD 296
Db 223 TL-----TVSKNATSHDLVLAAGGFIAIIDLVTGDTVALADALSRAKGIYQVGLFEG 278

```

```

Qy 297 H---PIGFPM-SLTYOEKATGSMCTVAKDFEYVKALEDGLISLDKARKMITGKYLK 352
Db 279 EFVVPATAMALKGLTLQGNVVG-----TYEVEVEVELARQG--SLPKL--PITGGLTV 329
Qy 353 DGEVKGFKOL 362
Db 330 DGVNDGLERL 339

```

```

Qy 297 H---PIGFPM-SLTYOEKATGSMCTVAKDFEYVKALEDGLISLDKARKMITGKYLK 352
Db 279 EFVVPATAMALKGLTLQGNVVG-----TYEVEVEVELARQG--SLPKL--PITGGLTV 329
Qy 353 DGEVKGFKOL 362
Db 330 DGVNDGLERL 339

```

```

Query Match 11.6%; Score 233.5; DB 4; Length 374;
Best Local Similarity 25.3%; Pred. No. 5,2e-16;
Matches 96; Conservative 57; Mismatches 142; Indels 85; Gaps 17;
Qy 17 TYPEPEIKRP--NDYKIKVSYGICGTDLKEFTYSGGVFPFKQGTKD-KISGYELPLC 74
Db 23 STEIEVAPPKTKVRIKILATICTRD--DHVIKTMV-----SKFPIYVG 67
Qy 75 HEFSGTVEVCGSVTSKPGDRVAVENTSHCSDRSRYKDTVAODLGICMACQSGSPNCA 134
Db 68 HEATGIVESIGEGVTIVKPDKVYLPFLPCRE-----CNCRNPDNLCI 113
Qy 135 SLSEFCGASG-----FAEVVYGEDHMKLPDSIPDDIGALVE 175
Db 114 RSDITRGVADGDTTFTCKGKPVHFMNTSTFETTVDESSVAKIDDAPEKVCILG 173
Qy 176 -PISVAM-HAVERARFQPGOTALVILGGPRLGATLILALOGHNAGKIYCSSEPALIRROFAK 233
Db 174 CGFSTGKGAAYKTKGKPGSTCYVFGGLGSLVINGCKSAGSRIIIGD--LNKDFEK 231
Qy 234 EL---GAEVFDP-STCDANAVLKAMPENEGFHAAFDCSGVPQFTTSIVATGPS-GIA 288
Db 232 AMAVGATECISPKDSTKPISEVLEMTGNVGY--TFEVIIGHLETMIDALASCHMNVGTS 289
Qy 289 VNVAVMGDHPICGMPMS---LTYOEKATGSMCTVAKDFEYVKALEDGLISLDKARKMI 345
Db 290 VVVGV-----PPSARKMLTYDP-----MLLETGRTWKGVFG--GLKSHDVPKLV 332
Qy 346 TGRVHLKDGVEKFKOLIEH 365
Db 333 TEFLAKKFDLD-----QLITH 348

```

Search completed: May 2, 2003, 13:05:55
 Job time : 240 secs


```

: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zysek, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 11975
: LENGTH: 352
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
: US-09-815-242-11975

Query Match      22.6% Score 455.5; DB 10; Length 352;
Best Local Similarity 33.5%; Pred. No. 3.9e-34;
Matches 127; Conservative 55; Mismatches 162; Indels 35; Gaps 12;

QY 1 MGGLVYGTNDIRSETVEPEIKNPNDKIKVSYGICGTDLKEFTYSGGVFPKQGT 60
DB 1 MALARHARARDLRLSEL--ERQAPRGEVELEAVAGCIGSDLHE--YQSGHSIP-QAE 55
QY 61 KDKISGELPLCPGHEFSGTVVEGCVTSVKRPGDRVAEATSHCSDSRSRYKDTVA 120
DB 56 AHPLESCRAPLTLGHEFCVVAALGPGVEGPRIGDRVAVEPEYRC-----G 101
QY 121 LCMACGSGSPNCASLSTFCGLGASGFAEYVYVYGEDHWKLPDIPDDIGALVPE 180
DB 102 ECHYCEGRNLCESMGFTGLMG--DGFARARAVPAYMLHRLPDVAGFQAALVLEPA 160
QY 181 WHAVERARPOPGOTALVLVGGPGLATILALOGHHAGKIVCSEPALIRROFAKE 239
DB 161 LHLRRSSILAPGRCACAVFGIGLILMLARLGIEDIAVADVSPERLALGEGFAS 220
QY 240 FPDSTDDANAVILKAVPENEGHAFDCSGVPGTFTTSIVATGPGCIANVAVMD 299
DB 221 LQARDDDTA---ARLRGGLALDCAFEAAGSGLDALASLRKGGELVLSLME--V 273
QY 300 GFPMNSLTQEKYATGSMCTYVD--FOEVVKALEDGLISLDRKARKITGVHLK 358
DB 274 RLDAFLVLRRELRLGSLVGC--RDATPELILALLADGRDLAR---VTSVPLE 328
QY 359 FKOLIEHKENKILVTPN 377
DB 329 FEALLRDK--SOLKVLVNP 346

RESULT 5
US-09-908-744-53
: Sequence 53, Application US/09908744
: Publication No. US20030068791A1
```

```

: GENERAL INFORMATION:
: APPLICANT: Misanikov Andrei
: APPLICANT: Ojamo Heikki
: APPLICANT: Povelainen Milla
: APPLICANT: Gros H+kan
: APPLICANT: Toivari Meri
: APPLICANT: Richard Peter
: APPLICANT: Ruohonen Laura
: APPLICANT: Kolvuranta Karl
: APPLICANT: Londeborough John
: APPLICANT: Aristidou Aristos
: APPLICANT: Penttil, Merja
: APPLICANT: Plazanet-Menut Claire
: APPLICANT: Deutscher Josef
: TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
: FILE REFERENCE: 1427.0010005
: CURRENT APPLICATION NUMBER: US/09/908,744
: CURRENT FILING DATE: 2001-07-20
: PRIOR APPLICATION NUMBER: PCT/FI01/00051
: PRIOR FILING DATE: 2001-01-22
: PRIOR APPLICATION NUMBER: US 09/488,581
: PRIOR FILING DATE: 2000-01-21
: PRIOR APPLICATION NUMBER: US 08/790,585
: PRIOR FILING DATE: 1997-01-29
: PRIOR APPLICATION NUMBER: US 08/368,395
: PRIOR FILING DATE: 1995-01-03
: PRIOR APPLICATION NUMBER: US 08/110,672
: PRIOR FILING DATE: 1993-08-24
: PRIOR APPLICATION NUMBER: US 07/973,325
: PRIOR FILING DATE: 1992-11-05
: NUMBER OF SEQ ID NOS: 70
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO: 53
: LENGTH: 350
: TYPE: PRT
: ORGANISM: Clostridium difficile
: US-09-908-744-53

Query Match      17.7% Score 357; DB 9; Length 350;
Best Local Similarity 28.9%; Pred. No. 5.5e-25;
Matches 111; Conservative 59; Mismatches 144; Indels 70; Gaps 12;

QY 1 MGGLVYGTNDIRSETVEPEIKNPNDKIKVSYGICGTDLKEFTYSG---GVFP 56
DB 1 MMSVRYGIRDRV--EDVVPKILEKDDVYIKKVAIGCIGSDISKSKTGPIMVGEIL-- 57
QY 57 KQTKDKISGELPLCPGHEFSGTVVEGCVTSVKRPGDRVAEATSHCSDSRSRYKDTVA 116
DB 58 -----GHEFSGEVAQVKEVASFKIGDRVAVCPAMPCE----- 91
QY 117 QDLGLCMACGSGSPNCASLSTFCGLGASGFAEYVYVYGEDHWKLPDIPDDIGALVPE 176
DB 92 -----CDECKKGLYSKCNVNAIIGKELGCGFAEYTKVERNLKIPDEISTETAAL 146
QY 177 ISVAMHARARPOPGOTALVLVGGPGLATILALOGHHAGKIVCSEPALIRROFAKE 236
DB 147 VCLAGHGLRSEAKVCDTVVLVGTPIGLFSLIOMAKIIFSTKLIANDVDEKDLAKELG 206
QY 237 AEYEDPSTDDANAVILKAVPE-----NEGHAFCGSGVPOTFTTSIVATGPGCI 290
DB 207 ADI-----C--INAKENIVEEIKRLTDGDADIVIESAGTPLTCQVILLAKKGT 259
QY 291 VANWGDHPIGFPMNSLTQEKYATGSMCTYV-----DFOEVVKALEDGLISL 340
DB 260 AGV---PYGDVALIREQEKIYRSEL--TVKCTWFGNSFPFGKMSAGLIYMOGD 313
QY 341 ARKMITGVHLKDG---VEGFKQ 361
DB 314 VEKLVTHRIINLEAPAYFEKVKR 337

RESULT 6
US-09-908-744-50
```

```

; Sequence 50, Application US/09908744
; Publication No. US20030068791A1
; GENERAL INFORMATION:
; APPLICANT: Miasnikov Andrei
; APPLICANT: Ojamo Heikki
; APPLICANT: Povelaainen Mira
; APPLICANT: Gros H+kan
; APPLICANT: Toivari Meri
; APPLICANT: Richard Peter
; APPLICANT: Ruohonen Laura
; APPLICANT: Koiivuranta Kari
; APPLICANT: Lonsborough John
; APPLICANT: Aristidou Aristos
; APPLICANT: Penttil, Merja
; APPLICANT: Plazanet-Ment Claire
; APPLICANT: Deutscher Josef
; TITLE OF INVENTION: Manufacture of Five-Carbon Sugars and Sugar Alcohols
; FILE REFERENCE: 1427.0010005
; CURRENT APPLICATION NUMBER: US/09/908,744
; PRIOR APPLICATION NUMBER: PCT/PT01/00051
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/488,581
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 08/790,585
; PRIOR FILING DATE: 1997-01-29
; PRIOR APPLICATION NUMBER: US 08/368,395
; PRIOR FILING DATE: 1995-01-03
; PRIOR APPLICATION NUMBER: US 08/110,672
; PRIOR FILING DATE: 1993-08-24
; PRIOR APPLICATION NUMBER: US 07/973,325
; PRIOR FILING DATE: 1992-11-05
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 50
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Bacillus halodurans
US-09-908-744-50

```

```

Query Match 17.5%; Score 352.5; DB 9; Length 354;
Best Local Similarity 27.6%; Pred. No. 1.5e-24;
Matches 108; Conservative 56; Mismatches 176; Indels 51; Gaps 9;

QY 1 MGKLLYGTNDIRSEVPEPEIKNPNDVYKISYCGICGTDLKEFTYSGGVFFPRQGT 60
DB 1 MKALNTYGIODLRPEET-PAPSTIEHDDIITIKKAVGICGSDLSRY----- 45
QY 61 KDRISGTELPICPGHEFSGTVEVSGVTSVKPGDRVAVEATSHCSRSRYKDTVAQDLG 120
DB 46 -KTLGPYVPMTGHEHAGEVYKIGRSVTFSGIDRYAACPTTTC-----G 90
QY 121 LCMACQSGSPNCASISFCGLGASGGFAEYVYGGEDHMKLPDIPDDIGALVEPTSA 180
DB 91 QCRCYCGEPTRCERLSVIG-ARHPGAYAEYVKLPKAKHVPDPNVMYDIALIEPASV 149
QY 181 WHAVERARPOGOTFALVGGSPIGLATTALOGHNAKTYCSEPALIRROFANELAEV 240
DB 150 AHGFTYTNIKPGASVAIMGSGISGLLAVOMAKIFGATYFAIDIDRQKLVANQLGADVL 209
QY 241 DPSTCDANAVLKAMPENEGFAAFDCSGVPOTFTTSIATGPGSIANVAVAGDHPIG 300
DB 210 ISSLQRAHAKQILEY-NGIGVUVAVESAGTSTSAQVAFALPKKGEEVFLGI---PYA 264
QY 301 FPMMSLTVEKRYATG-----SMCYTVKDFQEVVKALEDGLISLDKARKMITGKY 349
DB 265 DVOIEFFYEKIVRNELHYGSGWNLSPFGKEMATTIHYSSGGLNV---APMISTYKL 321
QY 350 HAKDVEKGFQKQILEKENNVKILVTPNEVS 380
DB 322 PLAKGPET-FOQIAKGLKPTKVLFEPEKIS 351

```

```

RESULT 7
US-09-815-242-10187
; Sequence 10187, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10187
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10187

```

```

Query Match 17.0%; Score 342; DB 10; Length 358;
Best Local Similarity 28.2%; Pred. No. 1.4e-23;
Matches 110; Conservative 66; Mismatches 162; Indels 52; Gaps 16;

QY 1 MGKLLYGTNDIRSEV-TYPEPEIKNPNDVYKISYCGICGTDLKEFTYSGGVFFPRQGT 59
DB 1 MKALARGKAFGKYMIDVDP-MCGPEYVIEIKAAALCGADMKHINVDG----- 51
QY 60 TKDRISGTELPICPGHEFSGTVEVSGVTSVKPGDRVAVEATSHCSRSRYKDTVAQDL 119
DB 52 -SDEFNSIR-----GHEFACICIAOGEKVKDMKVGQRYVSDNSGHV-----C 92
QY 120 GLCMACQSGSPNCASISFCGLGAS--GGFAEY-VYTG-----DHMYKLPDSIPDDI 170
DB 93 GVCPCAGEQGLCTCTEKVNLGDNNTWGGGFSKYCIVPGEILIKIRHNALEIPDGDYED 152
QY 171 GALVEPISYAMHAY-ERRARPOGOTFALVGGSPIGLATTALOGHNAKTYCSEPALIR 226
DB 153 AAVLDPICNAKRYKIAQGSFELPBGDDVYVIGTGLFVSOMARIMGAVNIWVGLQEDVA 212
QY 227 IRROFAKELGA-EVDPSTCDANAVLKAMPENEGFAAFDCSGVPOTFTTSIATGPGS 285
DB 213 VREPVAKELGATFVNGSTEDVVARQOICGKDNGL--VIESSGANIMALKALIEMLRN 270
QY 286 GLAVNVAWGDHPHIGMPMSLTVEKRYATGSMCTYKDFQEVVKALEDGLISLDKARKMI 345
DB 271 GEVVRVGM-GFPRPLDPSINDITAMNKSIIIGHMAYSTSWRNAIRLLASGAI---KVKPMI 326
QY 346 TGVKHLKDGVEKGFQKQILEKENNVKILVTPNEVS 375
DB 327 THRIGLSQWRE-GFDAMVD--KTAIKVIMT 353

```


[illegible]

Query Match	16.8%	Score 337.5;	DB 9;	Length 350;
Best Local Similarity	27.7%;	Pred. No. 3.6e-23;		
Matches 112;	Conservative 67;	Mismatches 144;	Indels 81;	Gaps 15;
QY	1	MKGLLYGTNDIYSETVEPEIK--NPNDVKIKVSYGICGTD----	LKEFTYSGPWF	54
Db	1	IKNAAVLHGINDMFEDI-----EIKPCEDEVKIKMAAGICGSDPPVYLKHWY---PV-	52	
QY	55	FPKQGTCKDKISGYELPLCGHEFGSGTVVSGVTSYGRDRAVVEATSHCSRSRYKDT	114	
Db	53	-----PAIRGHEFGSVIAEYKADYKNKVKYGRVVALYFICNE-----	90	
QY	115	VAQDDELGMACGSGSNCCASLSFCGLGASAGSRAEYVUGEDHMVKLPDSIPDICALV	174	
Db	91	-----CEYCKRGLESLCDHDMGLG-AKSGAEGAEVYNIKATNVLPITGMDPED--AAI	141	
QY	175	EPISVAMHVAVERAPQGTALVLLGGPIGIATLALOGHNHAKGYVSEPALRRQFAKE	234	
Db	142	EPILVAHMGVLNIGVQVDYVAVMMSGMGOLVLOGIKAGAGITIIAVDSDKMLRESKE	201	
QY	235	LGAEVDPSTCDANAAYLKAM-VPEBGEFHAAFDCSGVPOTFTTSIVATGPGSIAVNAV	293	
Db	202	LGADII--INADINPYRKIELGTGKGVDIALEACASKITQOCLLITKKKS-----	252	
QY	294	MGDHPGFEMPMS--LYTQEK-----YATSMCYTKYKDOEYVAKALEDL	335	
Db	253	-----KIGFLGIVSYITLSEAFENIRKLELKGFMNYSAPPGQEWTKGINLVNECK	308	
QY	336	ISLDARKMVTGKVAHLKDVEKGFROLIEHKENNVKLTLYPNEV	379	


```

Qy 241 DPSTCDANAVLKAMPENEGFHAAPDCSGVPOTFTTSIVATGSGIAVNAVWGDHPIG 300
Db 241 DPSTCDANAVLKAMPENEGFHAAPDCSGVPOTFTTSIVATGSGIAVNAVWGDHPIG 300
Qy 301 FPMPSLTYOEKRYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGVHLKDGVEKGF 360
Db 301 FPMPSLTYOEKRYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGVHLKDGVEKGF 360
Qy 361 QLEHKEKNVILVTPNPS 380
Db 361 QLEHKEKNVILVTPNPS 380

```

```

RESULT 2
US-10-147-003-2
; Sequence 2, Application US/10147003
; Publication No. US20030032153A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Kimoto, No. US20030032153A1, Hiro
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE,
; TITLE OF INVENTION: METHODS FOR PRODUCING SAME, AND METHODS FOR PRODUCING
; FILE REFERENCE: 06501-109001
; CURRENT APPLICATION NUMBER: US/10/147,003
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: JP 2001-159647
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Kluyveromyces fragilis
US-10-147-003-2

```

```

Query Match          52.9%  Score 1065.5; DB 9; Length 385;
Best Local Similarity 55.0%  Pred. No. 9.6e-91;
Matches 208; Conservative 56; Mismatches 111; Indels 3; Gaps 2;

Qy 1 MGCLLYGTNDIRYSETPPEIKNPNDYKIKVSYCGTDLKEFTYSGGVPFFPKGT 60
Db 1 MRALAYFGKODIRYTKDLPEYLETDDGEIEVSMCGISDLHE--YLDGPIFFEDCK 58
Qy 61 KDKISYELPLCPGHEFGSTVVEVSGVTSVKRPGDRVAVENTSHCSDRSRKYDTVAQDLG 120
Db 59 VHDVSLGLPQAWGHEMSISVSKVGPKNIRAKGDHVVVEATGTCLDHYTPNAAHAKDA 118
Qy 121 LCMACQSGSPNCASLSFPGGLGASGFAEYVYVYGEDHVMKLPDSIPDDIGALVEPISVA 180
Db 119 ECMAQGRGVYNCANHLGFGGLGVHSGFPEKVVYSEKHVVKLPNTLPDLVALVEPISVS 178
Qy 181 WHAVEARAFQCGTALVYGGPGLATLIALOGHHAGKIVCSEPALIRROFAKEIGAEVF 240
Db 179 WHAVRISKLQKOSALVYLCAGPGLATLIALOGHAGKIVCSEPALIRROFAKEIGAEVF 238
Qy 241 DPST-CDANAVLKAMPENEGFHAAPDCSGVPOTFTTSIVATGSGIAVNAVWGDHPI 299
Db 239 DPSEHKEDAVNLKILAPGEGFDRAYDCSGVKTFTDGVNHTTGRGVNVAIANGKPI 298
Qy 300 GPMPSLTYOEKRYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGVHLKDGVEKGF 359
Db 299 DKRPMDVTLQEKFTVGSMTYTIKDFEDVQALNGNSIADKRAHLITGRQIKEDGTGKF 358
Qy 360 KOLIEHKEKNVILVTPN 377
Db 359 DELMNHKEKNIKILLTPN 376

```

```

RESULT 3
US-09-815-242-11979
; Sequence 11979, Application US/09815242
; Patent No. US20020061569A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11979
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11979

```

```

Query Match          28.2%  Score 567.5; DB 10; Length 363;
Best Local Similarity 36.1%  Pred. No. 1.6e-44;
Matches 136; Conservative 73; Mismatches 141; Indels 27; Gaps 11;

Qy 1 MGCLLYGTNDIRYSETPPEIKNPNDYKIKVSYCGTDLKEFTYSGGVPFFPKGT 60
Db 11 MRALAYFGKODIRY-EDVPLPAEPFGVQIRVHMGCGISGDLHE--YLAQPVFIPE-A 66
Qy 61 KDKISYELPLCPGHEFGSTVVEVSGVTSVKRPGDRVAVENTSHCSDRSRKYDTVAQDLG 120
Db 67 PHLTGLKDCCLIGHSEGEIVRLGNGVTFRAVGOVAADACQHC-----G 112
Qy 121 LCMACQSGSPNCASLSFPGGLGASGFAEYVYVYGEDHVMKLPDSIPDDIGALVEPISVA 180
Db 113 TCYVCRHGLYNICEBLAFTGLMN-NGAFAEYVNPANLLVLPAGFSEAGALIEPLAVG 171
Qy 181 WHAVEARAFQCGTALVYGGPGLATLIALOGHHAGKIVCSEPALIRROFAKEIGAEVF 239
Db 172 WHAVKAKASGLIGQVNVVYVAGTIGLSTIMCARAGAAQVIALLEWSSARKAKALEVGAQV 231
Qy 240 FDPSTCDANAVLKAMPENEGFHAAPDCSGVPOTFTTSIVATGSGIAVNAVWGDHPI 299
Db 232 LDPSPRC-DALGETIALT-GGIGADVSEFCIGNKHTAKLALDAIKKACVYLVGTF-EPSP 288
Qy 300 GPMPSLTYOEKRYATGSMCYTVKDFQEVVKALEDGLISLDKARKMITGVHLKDGVEKGF 359
Db 289 EFNFEELVSTEEKOLLGLALVN-GEFADVIAFIADGRDI---APLVYGRIGLEIYBERGF 344
Qy 360 KOLIEHKEKNVILVTPN 376
Db 345 EELVNNKEKNVILVTPN 361

```

```

RESULT 4
US-09-815-242-11975
; Sequence 11975, Application US/09815242
; Patent No. US20020061569A1

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 2, 2003, 12:51:13 ; Search time 1361 Seconds
(Without alignments)
24.092 Million cell updates/sec

Title: US-10-020-674-2

Perfect score: 2013

Sequence: 1 MGGLYYGNDIRYSETYPE.....QLIEHKENKILLVPEVNS 380

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*\n2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep:*\n3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*\n4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*\n5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*\n6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*\n7: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep:*\n8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*\n9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*\n10: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*\n11: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*\n12: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*\n13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*\n14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2013	100.0	380	9	US-10-020-674-2
2	1065.5	52.9	385	9	US-10-147-003-2
3	567.5	28.2	363	10	US-09-815-242-11979
4	455.5	22.6	352	10	US-09-815-242-11975
5	357	17.7	350	9	US-09-908-744-53
6	352.5	17.0	354	9	US-09-908-744-50
7	342	17.0	358	10	US-09-815-242-10187
8	337.5	16.8	343	9	US-09-908-744-70
9	337.5	16.8	350	9	US-09-908-744-51
10	336	16.7	343	10	US-09-922-501-15
11	334	16.6	352	9	US-09-908-744-52
12	326.5	16.2	356	10	US-09-925-300-1326
13	319	15.8	343	10	US-09-815-242-5705
14	319	15.8	354	10	US-09-815-242-12464
15	315	15.6	347	10	US-09-815-242-13773
16	314	15.6	352	9	US-09-908-744-69
17	308	15.3	341	10	US-09-815-242-13833
18	307.5	15.3	346	10	US-09-741-669-421
19	307.5	15.3	346	10	US-09-815-242-10206

20	306	15.2	342	10	US-09-815-242-10958	Sequence 10958, A
21	284.5	14.1	457	9	US-10-002-245-4	Sequence 4, Appl1
22	282.5	14.0	455	10	US-09-775-009-7	Sequence 7, Appl1
23	277	13.8	347	10	US-09-815-242-12466	Sequence 12466, A
24	268.5	13.3	305	10	US-09-815-242-5706	Sequence 5706, Ap
25	265.5	13.2	343	10	US-09-741-669-412	Sequence 412, Ap
26	264	13.1	340	10	US-09-971-361-10	Sequence 10, Appl
27	259	12.9	354	9	US-10-166-087-6	Sequence 6, Appl1
28	255	12.7	369	10	US-09-854-132-40	Sequence 40, Appl
29	252.5	12.5	359	10	US-09-815-242-13653	Sequence 13653, A
30	246	12.2	346	10	US-09-815-242-10162	Sequence 10162, A
31	239.5	11.9	336	10	US-09-815-242-12430	Sequence 12430, A
32	239.5	11.9	339	10	US-09-815-242-5496	Sequence 5496, Ap
33	236	11.7	352	10	US-09-954-314-18	Sequence 18, Appl
34	231	11.5	336	10	US-09-815-242-13832	Sequence 13832, A
35	230.5	11.5	349	9	US-09-908-744-49	Sequence 49, Appl
36	228.5	11.4	342	10	US-09-815-242-12104	Sequence 12104, A
37	228	11.3	368	9	US-09-738-626-3857	Sequence 3857, Ap
38	225.5	11.2	375	9	US-09-981-353-113	Sequence 113, App
39	225.5	11.2	386	9	US-09-854-133-81	Sequence 81, Appl
40	225.5	11.2	386	10	US-09-738-973-81	Sequence 81, Appl
41	222	11.0	357	10	US-09-947-027-9	Sequence 9, Appl1
42	222	11.0	357	12	US-10-091-009-9	Sequence 9, Appl1
43	220.5	11.0	264	10	US-09-815-242-11710	Sequence 11710, A
44	220	10.9	362	10	US-09-947-150-2	Sequence 2, Appl1
45	220	10.9	362	10	US-09-947-027-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1\nUS-10-020-674-2\n: Sequence 2, Application US/10020674\n: Patent No. US20020160468A1\n: GENERAL INFORMATION:\n: APPLICANT: Yamamoto, Hiroaki\n: APPLICANT: Onodera, Keiko\n: APPLICANT: Tani, Yoshiki\n: TITLE OR INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE\n: FILE REFERENCE: 06501-092001\n: CURRENT APPLICATION NUMBER: US/10/020, 674\n: CURRENT FILING DATE: 2002-04-19\n: PRIOR APPLICATION NUMBER: JP 2000-33363\n: PRIOR FILING DATE: 2000-10-31\n: NUMBER OF SEQ ID NOS: 17\n: SOFTWARE: FastSeq for Windows Version 4.0\n: SEQ ID NO 2\n: LENGTH: 380\n: TYPE: PRT\n: ORGANISM: Pichia angusta\n: US-10-020-674-2

Query Match	100.0%	Score 2013:	DB 9:	Length 380:
Best Local Similarity	100.0%	Pred. No. 8.9e-179:		
Matches 380:	Conservative	0:	Mismatches	0:
			Indels	0:
			Gaps	0:
QY	1	MGGLYYGNDIRYSETYPEPEIKNPNDVKIKVSCGICGDKLKEFTYSGGVFPFKGT	60	
DB	1	MGGLYYGNDIRYSETYPEPEIKNPNDVKIKVSCGICGDKLKEFTYSGGVFPFKGT	60	
QY	61	KDKISGYELPLCPGHEFSGTVVGVGSYTSVAKPGRAVAVATSHCSDSKSRKDYTAODLG	120	
DB	61	KDKISGYELPLCPGHEFSGTVVGVGSYTSVAKPGRAVAVATSHCSDSKSRKDYTAODLG	120	
QY	61	KDKISGYELPLCPGHEFSGTVVGVGSYTSVAKPGRAVAVATSHCSDSKSRKDYTAODLG	120	
DB	61	KDKISGYELPLCPGHEFSGTVVGVGSYTSVAKPGRAVAVATSHCSDSKSRKDYTAODLG	120	
QY	121	LCMACQSSPNCASISFCGIGAGSGFAEYVYGEDHVKLPDLSIPDICALVEPISVA	180	
DB	121	LCMACQSSPNCASISFCGIGAGSGFAEYVYGEDHVKLPDLSIPDICALVEPISVA	180	
QY	121	LCMACQSSPNCASISFCGIGAGSGFAEYVYGEDHVKLPDLSIPDICALVEPISVA	180	
DB	121	LCMACQSSPNCASISFCGIGAGSGFAEYVYGEDHVKLPDLSIPDICALVEPISVA	180	
QY	181	WHAVERARQPOQTALVLGGGPIGLATILALOGHNAGKIVCEPALIRROFAKEIGAEYF	240	
DB	181	WHAVERARQPOQTALVLGGGPIGLATILALOGHNAGKIVCEPALIRROFAKEIGAEYF	240	

RESULT 14
 US-09-815-242-12464
 : Sequence 12464, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlssen, Karl L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : FILE OF INVENTION: ELITRA.01A
 : FILE REFERENCE: ELITRA.01A
 : CURRENT APPLICATION NUMBER: US/09/815,242
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR APPLICATION NUMBER: 60/191,078
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR APPLICATION NUMBER: 60/206,848
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR APPLICATION NUMBER: 60/207,727
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR APPLICATION NUMBER: 60/242,578
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR APPLICATION NUMBER: 60/253,625
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR APPLICATION NUMBER: 60/257,931
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR APPLICATION NUMBER: 60/269,308
 : PRIOR FILING DATE: 2001-02-16

Search completed: May 2, 2003, 15:32:41
Job time : 1363 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 07:40:56 ; Search time 3103 Seconds
(without alignments)
10720.112 Million cell updates/sec

Title: US-10-020-674-1

Perfect score: 1143

Sequence: 1 atgaaggcttacttata.....cgccgaacgaggttctctaa 1143

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_hlg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_pl:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hlg_hum:*

31: em_hlg_inv:*

32: em_hlg_other:*

33: em_hlg_mus:*

34: em_hlg_pln:*

35: em_hlg_rtd:*

36: em_hlg_mam:*

37: em_hlg_vrt:*

38: em_sy:*

39: em_higo_hum:*

40: em_higo_mus:*

41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	274	24.0	103682	8	SCU12980	U12980 Saccharomyc
2	163	14.3	915	11	CNS06104	AL339674 73 end of
3	83	7.3	1041	6	AX366887	AX366887 Sequence
4	83	7.3	16585	1	AB007638	AB007638 Bacillus
5	83	7.3	213190	1	BSU09004	299107 Bacillus su
6	69.8	6.1	339	6	AX312470	AX312470 Sequence
7	67.4	5.9	717	11	CNS0686L	AL402499 77 end of
8	67	5.9	1126	8	AV133848	AV133848 Arabidops
9	67	5.9	1269	8	AF370161	AF370161 Arabidops
10	67	5.9	1347	8	AY085213	AY085213 Arabidops
11	65.8	5.8	1457	8	AB025969	AB025969 prunus pe
12	64.4	5.6	10582	1	AE008211	AE008211 Agrobacte
13	64.4	5.6	11742	1	AE009402	AE009402 Agrobacte
14	64.2	5.6	81347	8	AB015478	AB015478 Arabidops
15	63.8	5.5	15088	1	AE004831	AE004831 Pseudomon
16	62.8	5.5	1404	3	AY052067	AY052067 Drosophila
17	62.2	5.4	1380	3	AY058731	AY058731 Drosophila
18	61.4	5.4	1456	8	AY037946	AY037946 Prunus ce
19	60.8	5.3	11248	1	AE006323	AE006323 Lactococc
20	59.6	5.2	1526	8	SPTWS1	X74422 S. pombe tms
21	59.6	5.2	40438	8	SPU41410	U41410 Schizosacch
22	59.6	5.2	40810	8	SPBC1773	AL033389 S. pombe c
23	59.6	5.2	305153	1	AP001520	AP001520 Bacillus
24	58.6	5.1	1572	8	AB042810	AB042810 E. coli
25	58.6	5.1	25025	3	CER0485	Z70782 Caenorhabdl
26	57.4	5.0	933	6	ARI65148	ARI65148 Sequence
27	57	5.0	1396	8	AF323504	AF323504 Malus x d
28	57	5.0	249050	1	AB016256	AB016256 Malus dom.
29	56.6	5.0	332635	1	AL596165	AL596165 Listeria
30	56.6	5.0	349980	6	AP003005	AP003005 Mesorhizo
31	56.6	5.0	349980	6	AX417038	AX417038 Sequence
32	56.6	5.0	349980	6	AX417041	AX417041 Sequence
33	55.8	4.9	24336	1	AE008841	AE008841 Salmonell
34	55.4	4.8	1038	6	AX377790	AX377790 Sequence
35	52.8	4.6	2921	3	DMAF02212	AF002212 Drosophila
36	52.8	4.6	72847	3	AC004266	AC004266 Drosophila
37	52.8	4.6	110000	3	AE001572_0	AE001572 Drosophila
38	52.8	4.6	152108	2	AC020282	AC020282 Drosophila
39	52.8	4.6	165761	3	AC095016	AC095016 Drosophila
40	52.8	4.6	174140	3	AC008225	AC008225 Drosophila
41	52.8	4.6	309357	3	AE003673	AE003673 Drosophila
42	52.2	4.6	1808	9	HSU07361	U07361 Homo sapien
43	52.2	4.6	2487	9	BC025295	BC025295 Homo sapi
44	52.2	4.6	2519	6	AX330493	AX330493 Sequence
45	52.2	4.6	2519	6	AX337243	AX337243 Sequence

ALIGNMENTS

RESULT 1

SCU12980

LOCUS SCU12980 103682 bp DNA linear PLN 05-MAR-1998

DEFINITION Saccharomycetes cerevisiae chromosome I left arm sequence.

ACCESSION U12980 U00091

VERSION U12980.1 GI:2911250

KEYWORDS S.cerevisiae ycr128p homology; FTO9; GDN3; two alcohol/sorbitol dehydrogenase homologs; SIM1; CNE1; ACS1; S.pombe SPAC 1F7.03 homology; S.cerevisiae Pip2p-like transcription factor homology; GCV3; glycine cleavage H protein; PTA1; FUN9 transcript; essential gene; S.pombe SPAC 24B11.08c homology; CDC24; CUN3/WHI1/DAF1; CYC3; PYK1; S.cerevisiae ORF 06283 homology; Xenopus laevis GTP-binding protein DRG homology; FUN12 transcript; essential gene; FUN19 transcript; S.cerevisiae ori 06265 homology; FUN33 transcript; essential gene; FUN20 transcript; essential gene; S.pombe SPAC

SOURCE ORGANISM	BAA 06 homolog; SNCl, MYO2; S. cerevisiae ORF 06159 homolog; DRS2; MAK16; LTEL.
REFERENCE AUTHORS TITLE	Saccharomyces cerevisiae. Saccharomyces cerevisiae Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
JOURNAL MEDLINE PUBMED	Nagasu,T. and Hall,B.D. Nucleotide sequence of the GDH gene coding for the NADP-specific glutamate dehydrogenase of Saccharomyces cerevisiae Gene 37 (1-3), 247-253 (1985)
REFERENCE AUTHORS TITLE	EMBO J. 6 (1), 235-241 (1987) 2932370
JOURNAL MEDLINE PUBMED	Dumont,M.E., Ernst,J.F., Hampsey,D.M. and Sherman,F. Identification and sequence of the gene encoding cytochrome c heme lyase in the yeast Saccharomyces cerevisiae EMBO J. 6 (1), 235-241 (1987)
REFERENCE AUTHORS TITLE	87218469
JOURNAL MEDLINE PUBMED	3 (sites) Miyamoto,S., Ohya,Y., Ohsuni,Y. and Anraku,Y. Nucleotide sequence of the Cls4 (CDC24) gene of Saccharomyces cerevisiae Gene 54 (1), 125-132 (1987)
REFERENCE AUTHORS TITLE	87277425 3301539
JOURNAL MEDLINE PUBMED	4 (sites) Wickner,R.B. Host function of MAK16: G1 arrest by a mak16 mutant of Saccharomyces cerevisiae Proc. Natl. Acad. Sci. U.S.A. 85 (16), 6007-6011 (1988)
REFERENCE AUTHORS TITLE	88320371 3045810
JOURNAL MEDLINE PUBMED	5 (sites) Nash,R., Tokiwa,G., Anand,S., Erickson,K. and Fletcher,A.B. The WHI1+ gene of Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog EMBO J. 7 (13), 4335-4346 (1988)
REFERENCE AUTHORS TITLE	89210821 2907481
JOURNAL MEDLINE PUBMED	6 (sites) McNally,T., Purvis,I.J., Fothergill-Gilmore,L.A. and Brown,A.J. The yeast pyruvate kinase gene does not contain a string of non-preferred codons: revised nucleotide sequence FEBS Lett. 247 (2), 312-316 (1989)
REFERENCE AUTHORS TITLE	89232143 2653861
JOURNAL MEDLINE PUBMED	7 (sites) Harris,S.D., Cheng,J., Pugh,T.A. and Pringle,J.R. Molecular analysis of Saccharomyces cerevisiae Chromosome I. On the number of genes and the identification of essential genes using temperature-sensitive-lethal mutations J. Mol. Biol. 225 (1), 53-65 (1992)
REFERENCE AUTHORS TITLE	92260538 1583694
JOURNAL MEDLINE PUBMED	8 (sites) Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandrakl,D., Antoline,G., Anwar,R., Ballesta,J.P., Benit,P. et al. The complete DNA sequence of yeast chromosome III Nature 357 (6373), 38-46 (1992)
REFERENCE AUTHORS TITLE	92244356 1574125
JOURNAL MEDLINE PUBMED	9 (sites) Gest,J.E., Rodgers,L., Riggs,M. and Wiegler,M. SNCl, a yeast homolog of the synaptic vesicle-associated membrane protein/synaptobrevin gene family: genetic interactions with the Ras and CAP genes Proc. Natl. Acad. Sci. U.S.A. 89 (10), 4338-4342 (1992)
REFERENCE AUTHORS TITLE	92262435 1316605
JOURNAL MEDLINE PUBMED	10 (sites)
REFERENCE	10 (sites)
AUTHORS TITLE	O'Connor,J.P. and Peebles,C.L. PTA1, an essential gene of Saccharomyces cerevisiae affecting pre-tRNA processing Mol. Cell. Biol. 12 (9), 3843-3856 (1992)
JOURNAL MEDLINE PUBMED	92375052 1508188
REFERENCE AUTHORS TITLE	11 (sites) De Virgiliio,C., Burckert,N., Barth,G., Neuhaus,J.M., Bolter,T. and Wiemken,A. Cloning and disruption of a gene required for growth on acetate but not on ethanol: the acetyl-coenzyme A synthetase gene of Saccharomyces cerevisiae Yeast 8 (12), 1043-1051 (1992)
JOURNAL MEDLINE PUBMED	93190633 1363452
REFERENCE AUTHORS TITLE	12 (sites) de Virgiliio,C., Burckert,N., Neuhaus,J.M., Bolter,T. and Wiemken,A. CNEI, a Saccharomyces cerevisiae homologue of the genes encoding mammalian calnexin and calreticulin Yeast 9 (2), 185-188 (1993)
JOURNAL MEDLINE PUBMED	93220396 8463605
REFERENCE AUTHORS TITLE	13 (sites) Teunissen,A.W., Holub,E., van der Hucht,J., van den Berg,J.A. and Steensma,H.Y. Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisiae Yeast 9 (4), 423-427 (1993)
JOURNAL MEDLINE PUBMED	93289821 8511970
REFERENCE AUTHORS TITLE	14 (sites) Ripmaster,T.L., Vaughn,G.P. and Woolford,J.L. Jr. DRS1 to DRS7, novel genes required for ribosome assembly and function in Saccharomyces cerevisiae Mol. Cell. Biol. 13 (12), 7901-7912 (1993)
JOURNAL MEDLINE PUBMED	94067151 8247005
REFERENCE AUTHORS TITLE	15 (bases 99096 to 103271) Keng,T., Clark,M.W., Storme,R.K., Fortin,N., Zhong,W., Ouelllette,B.F., Barton,A.B., Kaback,D.B. and Bussey,H. LTEL of Saccharomyces cerevisiae is a 1435 codon open reading frame that has sequence similarities to guanine nucleotide releasing factors Yeast 10 (7), 953-958 (1994)
JOURNAL MEDLINE PUBMED	95076714 7985422
REFERENCE AUTHORS TITLE	16 (sites) Suttrave,P., Shafer,B., Strathern,J. and Hughes,S. Isolation, identification and characterisation of the FUN12 gene of Saccharomyces cerevisiae Gene (1994) In press
JOURNAL MEDLINE PUBMED	17 (bases 1 to 103682) Bussey,H., Kaback,D.B., Zhong,W., Vo,D.T., Clark,M.W., Fortin,N., Hall,J., Ouelllette,B.F.F., Keng,T., Barton,A.B., Su,Y., Davies,C.K. and Storme,R.K. The nucleotide sequence of chromosome I from Saccharomyces cerevisiae Proc. Natl. Acad. Sci. U.S.A. 92 (9), 3809-3813 (1995)
REFERENCE AUTHORS TITLE	95249563 7731988
JOURNAL MEDLINE PUBMED	18 (sites) Harer,B. and Brown,S. Identification of the yeast myosin gene that is similar to the yeast MYO2 gene Unpublished (1992)
REFERENCE AUTHORS TITLE	19 (bases 1 to 103682) Volckaert,G. and Valle,G. Unpublished (1997)
JOURNAL MEDLINE PUBMED	20 (bases 1 to 103682)
REFERENCE	Direct Submission

REFERENCE	TITLE	AUTHORS	REMARK	AUTHORS	TITLE	JOURNAL
JOURNAL	Submitted (04-ANC-1994)	Danh Vo, Biology, MCGill University, 1205				
REFERENCE	21 (bases 1 to 103682)	Docteur Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada				
JOURNAL	Vo,D.					
JOURNAL	Direct Submission					
REMARK	Submitted (10-APR-1996)	Danh Vo, Biology, MCGill University, 1205				
AUTHORS	Docteur Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada					
REFERENCE	Sequence update which affects many coding regions					
AUTHORS	22 (bases 1 to 103682)					
TITLE	Jia,Y. and Cherry,J.M.					
JOURNAL	Direct Submission					
	Submitted (27-OCT-1997)	Department of Genetics, Stanford				
	University, Saccharomyces Genome Database, Stanford, CA 94305-5120,					
	USA					
COMMENT	On Feb 25, 1998 this sequence version replaced gi:1326053.					
	Curated by:					
	Saccharomyces Genome Database					
	URL: http://genome-www.stanford.edu/					
	e-mail: yeast-curator@genome.stanford.edu					
FEATURES	Location/Qualifiers					
source	1..103682					
	/organism="Saccharomyces cerevisiae"					
	/strain="S288C/AB972"					
	/db_xref="taxon:4932"					
	/chromosome="I"					
	/map="1L: subtelomeric region to LTR1"					
	/complement(4645..6426)					
	/gene="SEO1"					
	/complement(4645..6426)					
	/gene="SEO1"					
	/note="YAL067C: Suppressor of sulfoxide ethionine					
	resistance; similar to S. cerevisiae Ycr28p Swiss-Pro					
	Accession Number P25621, with homology to allantoate					
	permease"					
	/citation="(8)					
	/codon_start=1					
	/product="Seo1p: putative membrane protein"					
	/protein_id="AAC04968.1"					
	/db_xref="GI:595520"					
	/translation="MYSIKREIIVDPYKRLKMGFIPVKRQVEDLPDDLNSREIYIS"					
gene						
CDS						
	Query Match 24.0%; Score 274; DB 8; Length 103682;					
	Best Local Similarity 53.9%; Pred. No. 5.1e-66;					
	Matches 612; Conservative 0; Mismatches 515; Indels 9; Gaps 2;					
1	ANGAAGGTTACTTATTACGCTACAAAGCATATTCGCTACTCGAAGAGGTTCTCTGA 60					
DB	32565 ATGAAAGCTTGGCATATTTCAGAAAGGCTGATATTCACTTACATGATATCCCTAG 32624					
61	CCGAGATCAAGATCCCAACGATCTCAAGATCAAGATCAAGTACGTTATGTTGGAATCTGTGC 120					
DB	32625 CGAAGATCCAAACCGAGCATAGGTTATATTCAGAGTCTCTGTGGTGGATTTGTGCG 32684					
121	ACGGAATTGAAGAAATTCATATTTCTGAGGCTCTGTTTTCCTTAACAAGGACACC 180					
DB	32685 TCGGATCTTCACGAGTACTTGGAT-----GGTCAATCTTCATGCTCAAGATGGAGAG 32738					
181	AAGGACAAGATTTCCGGGATACGAACACTCTCTCTGTCTGCTCGAGCATGAATTTAGCGGANG 240					
DB	32739 TGGCATTAATTTTCCAAAGCTCTTTACTCTGGGCAATGGGCAATGATGTCAGAAAT 32798					
241	GTGGTACGAGTTGGCTCTGTGTTCAAGTGTGAACCTGGTACAGAGATCGCAGTTGA 300					
DB	32799 GTTTCCAAAGTTGTGTCCTTAATGATGCAAGAGTGAAGTTGGGACACAGTGGTCTGAT 32858					
301	GCTACGTGCAATGCTCCGACAGATCGGCTCAAGAGCAGGTCGCCCAAGACCTTGGG 360					
DB	32859 GCTGCACAGAGTTGTGGGACCTCGATTGCTGGCCACACTCCAAATTTTCAATTCANA 32918					
361	CTCTGTATGGCTCCGACAGCGGATCTCCGAACCTGCTGTGGCTCGCTAGACTTTGCGGT 420					
DB	32919 CCATGTGATGCTTTCAGAGGGGACGTGAATAATCTATGATACCCAGCCGCGTATTGTA 32978					

Accession	Version	Keywords	Source	Organism	Reference	Title	Journal
QY 421	1						
Db 32979	1						
QY 481	1						
Db 33039	1						
QY 541	1						
Db 33099	1						
QY 601	1						
Db 33159	1						
QY 661	1						
Db 33219	1						
QY 721	1						
Db 33279	1						
QY 778	1						
Db 33339	1						
QY 838	1						
Db 33399	1						
QY 898	1						
Db 33459	1						
QY 958	1						
Db 33519	1						
QY 1018	1						
Db 33579	1						
QY 1078	1						
Db 33639	1						
RESULT 2							
CNS06104/C							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
AUTHORS							

Db	458	CTTCGACGTTGCTCTAATAGCGTCTCCGCTCAAGCAAACTCAAAAGCGAGGACAAAGCGG	517
Oy	587	TGCTTTCTTGGAGAGAGTTCATAGCGCCTTGCCACACATTCTTGCTCTGCAAGCCATCATG	646
Db	518	CTGTATTTCGGGTGGGGCCCGATCGCATTTCTTGATTTGAAGCGCTGAAGCGTCCGGTG	577
Oy	647	CGGSCAAAATGTGTGTTCCGAGCCGGCCCTTGATCAGAAAGACGTTTGCAAGGAAGCTGG	706
Db	578	CAACTGATATTTCACGCTGTGAGCTTCTCTCTGAAGCCAGCAAAAAGCTGAGAGACTTG	637
Oy	707	GCCTGTAAGTGTTCGATCTCTTACATGTAGACGACGCAAAATGCTGTTCTCAAGGCTATGG	766
Db	638	GGCGGATCATCTGTGATCCGCTTAAACAAACGATGTAGTCGGTG-----AGATTGCAG	691
Oy	767	TGCCGAGAAACGAGGGATTCATCCAGCCCTTCGACTCGTGGGTTCCTCAGACAAATCA	826
Db	692	AACGTACAGAGAGCGCGGTGTGACGTACCATTTCCAGTCACTGGTGTCCACGTGTGTAC	751
Oy	827	CCACCTCAATTGTGCCACAGGGACCTTCTGGAATCGCCGTCAAATGTGGCCGTTTGGGAG	886
Db	752	GACAAGCCATCCACTCCACTCAATATTCGCGTGAACCGTCATCGTCAGCATTTGGGAA	811
Oy	887	A 887	
Db	812	A 812	

gene
1563..2541
/gene="guta"
93..1484
/note="putative"
/codon_start=1
/transl_table=1
/product="sugar transport protein"
/protein_id="BAA22759.1"
/db_xref="GI:2522008"
/translation="MIGTEKVTALKAKHQSSTGCLSNFERIGYFGEDMSNIIFQV
NAVLYEYTDVGGIOPAVIATFLVWRVLDIIFDPINGLIIDKNTNMGKARPLWV
AFPEFLFDFCTTDPHGEETNMVYAVVTYILDMSESMQIPVNSLTGRNYSER
TVLTTRNMLIVYIGILSLISCATPLATVIGEDQAFGCMALYAAVSYLVNLSFF
TVREIQKRRKKKQIKKTLISVLEFNKRLMLLISPLAFAIGFNKISTWYVTVVW
NHKEVFNVTGLFEGFALISNLIPLPFSEKGRQWMIITPALSISYAGIHFPFSS
IPLIFVFAAGFETPLNTLPMGVMDVCVYAPMKQIRADQVVISMSPIKGLGA
LAGSSALYGLGIAGVANTDOTVASLNKIKNNMLIFQFILLSIILAFPLTERKY
KHIISELQRSR"

gene
1563..2541
/gene="fruc"
1563..1574
/gene="fruc"
1579..2541
/gene="fruc"
/note="putative"
/codon_start=1

CDS

RESULT 4				
AB007638/c				
LOCUS	AB007638	16585 bp	DNA	linear
DEFINITION	Bacillus subtilis genomic DNA containing gula to coxA region, 48			
ACCESSION	AB007638			
VERSION	AB007638			
	AB007638.1	GI:2522006		

/protein_id="BAA22760.1"
 /db_xref="c1:2522009"
 /translation="MRSHSVVICGELLIDFCTDYVDNDLMGRGRLKLSAGCAPANVSA
 AIAKGGDAASGKVGKDPFGFLKRTIDAVHVDSTLMDEKAPPTLALFAFLSKNKE
 RDNVNRKADALFTLEDIDOKELNFKALHFGSAATLSDPFGASVLRNLASINDNO
 FISEDPNREDLNRGVSEFVSAKKAALVSDPKVSEDELEIIISGVKDHQGVALLH
 EIGAVIVATVLTGKSGTLLSGKDRRIIVTSIVTSDISGADAAVGAALVQALNTDII

REFERENCE AUTHORS TITLE	SOURCE ORGANISM
1 (sites)	<i>Bacillus subtilis</i> (strain:Marburg 160) DNA.
Kashimura, Y., Nakai, S., Ogasawara, N., Yata, K. and Sadate, Y.	<i>Bacillus subtilis</i>
Sequence analysis of the <i>groEL</i> - <i>cola</i> region of the <i>Bacillus</i>	Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
subtilis genome, containing the restriction/modification system	
genes	
DNA Res. (1997) In press	
2 (bases 1 to 16585)	
Sadate, Y.	
Journal	
Reference	
Authors	
Title	
Journal	
Submitted (03-Oct-1997)	
Yoshito Sadate, National Institute of	

```

gene      2722. .3420
          /gene="ydfF"
RBS       2722. .3277
          /gene="ydfF"
          /gene="ydfF"
CDS       2737. .3420
          /gene="ydfF"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /product="membrane protein"
          /protein_id="BAA2761.1"
          /db_xref="GI:2522010"
          /translation="MSIGRFKIMSANINALLDKAEINPEKMDVYLIRNNMSDLAKVK

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
DNA Res. (1997) In Press
2 (pages 1 to 1658)
Sadale, Y.
Direct Submission
Submitted (03-OCR-1997) Yoshiko Sadale, National Institute of
Genetics, Radioisotope Center, Yata 1111, Mishima, Shizuoka 411
Japan (E-mail: yosadale@lab.nig.ac.jp, Tel: 81-0559-81-6870,
Fax: 81-0539-81-6870)

```

/transl_table=1
/product="membrane protein"
/protein_id="BAA22761.1"
/db_xref="GI:2522010"
/translation="MSITGRFRDIMSANINALLDKAKENPEKMDVOYLRRNMSDLAKYK
AETAYVAHEQRAKREYHECCADMEKMSYAMKALQGNESDARKPLERTSLSEKLS
ELQANQIATNIAADMRKHDRLVSDIGELARKNMIAKAWAKTGRKNNLCAASV
STQSMQSFGRNEKDVYNALDOANNAEMELNSAPDDMDADLSAKYDTQSGSSQVYDELAA

```

FEATURES	Location/Qualifiers
source	1. .16585

gene LKAKMMLDK" join(3472. .3479,3486. .4511)

gene
CDS

```
/gene="ydg"  
/note="function unknown"  
/codon_start=1
```

```

/EC_number="1.1.1.14"
/note="L-Idiotol 2-dehydrogenase"
/codon_start=2
/transl_table=11
/product="sorbitol dehydrogenase"
/protein_id="BA22869.1"
db_xref="refseq:NM_001143270.1"

```

```

/protein_id="BAA22762.1"
/db_xref="GI:2522011"
/translation="MIIISKPCNCGSDMAFDSEITGSLSCSSCGRODNIESLIPKENTIAA
RPSDEAEAYECKNCGAVLITAEATTATGSCFGCAALLDRLSGHAPAVIPFTTS
KQEAQARMKCKGGLTPRGFSMDRKSTGTAYIPPMWDLSEVQNRNRYHQ
YEEGYYICTETIEHBEAFDINDILYLIKIPVDSSEKKMDELMKLEPYSTEEELKDQYAY

```

gene
RBS

```

'LMGCGF"
4497.4501
/gene="ydhH"
gene

```

```

BBS      4497..4501
          /gene="ydljH"
          /gene="ydljH"
          4511..5275
          /gene="ydljH"
          4511..5275
          /gene="ydljH"
          /note="function unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAA22763.1"
          /db_xref="GI:2522012"
          /translation="WRGFGKAIFVIAVLTMPDLGIEAVRASELQOHVYDRAHLIS
          KAEIGLESLKAKGAKRDPTDPIIITTKSTGDEIDAYGTGDFYDRYKRSFALITIDM
          NREVEIFGFGKAEQYIDNRSILNRSIRNTISDLNENFEAFETYIOLSDMDKIRKG
          VNPDIETFWMEFOLIAIIVAGGIAVSIMLYIAGKGVVNGSTYMDQSTSDIDQYDIY
          IRTVTRKRKSGSGSDGSGDYTKGTSYSSRSRS"
          5296..6277
          /gene="ydljI"
          5296..6301
          /gene="ydljI"
          5306..6277
          /gene="ydljI"
          /note="function unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAA22764.1"
          /db_xref="GI:2522013"
          /translation="MSFERNQIANVWEVEEERDDMIFYKNNREIKKGRSLITPGD
          AVFLNGENFGEVODGDYDIESEIPIELSLKGFKEGNGRAEVLVFNTRKEFTR
          WGTCAINIPAGMGDMPIRANGTFNPKVODYSLIDKIAGVDOYDFVEDIRITIS
          ILDDITIMNATIRREGDMPNLCQANPDIAKGQEDLDQILSDGTYIGFOIMSNYQ
          EVQDITITNASTYGMAGVDYNRQQOISMTDGNASGMSGSGASDAGMMGMNNAQNM
          NQNMNQDAQSGSGPOSTSGSKPNFCPCNGKTEANFCPCNGKLV"
          6291..6323
          /note="stem-loop"
          complement(6324..7365)
          /gene="ydljY"
          complement(6324..7349)
          /gene="ydljY"
          /note="function unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAA22765.1"
          /db_xref="GI:2522014"
          /translation="MNEGSIITAGLLITAGMIDPLMTIMESGAGPTKCSANLAK
          GCRISGHAUVLMSGPIILCTLYVWISLPSGWTLYISDPSHSLMETQKSPASW
          SDRIYFSCTYMTLGNGBLAPNGGLMKLVITITETAGLLITTESVTYLLISLVASVOK
          RSPQSVSITLSDHDTETELVHNAMNGKRDHDDIFLVAASSELGKTAQHNAPLITHEY
          STQDSESSILAVVAIDELTIFKGIPEQVQPNQILHTEARSSIKNTLDVHVAYIHP
          AEOAPPEPDISKLOOSGIPALSKOTFQIAVNSIKERQLLGLITQAGRRKVPQEOAI
          GNAVSPK"
          complement(7358..7365)
          /gene="ydljY"
          7922..9355
          /gene="ydljK"
          7922..7927
          /gene="ydljK"
          /gene="ydljK"
          7934..9355
          /gene="ydljK"
          /note="putative"
          /codon_start=1
          /transl_table=11
          /product="metabolite transport protein"
          /protein_id="BAA22766.1"
          /db_xref="GI:2522015"
          /translation="MNKQGNQMSFLRTIILVSTFGGILBEYDTGVNGLALPYGEPDQ
          LNLAFIEGLVTSILFGALGAVFGGMSDFNGRRNNILFLAVIFGISITGTFAPN
          VTWIAISRFVILAVGASVTPAYVIALEMSVPSRGMVQNEIMETIVSQGLAFVNA
          ILGTGQDSNVHWRFLVILASLPALELFEGMIRPESRMLVSKGRKEDLRYLKR
          DERKAAAELODIEFAEKEDOLEKATFSDVPMVRVITIGIAIVOOITGVNSIM
          YGEIIEINSGSEPTALIGNINGVATVATVGTWILGVRGRPMMLGLIGTGA
          LLIGTISIVESGEPALPYVIVSLTVFLFALQOQASIPYTWMLSETPPLRLGMDG
          VTVECLMMVNAVSFTFPIILAIIGLSITTFIVGAGICSVLTVKRRLPETKGLSIOQ

```

	terminator	LEENPRAVDHSHKAKDSAEVIG"
	gene	/note="stem-loop" complement(9403..10457) /gene="ydlJL" complement(9403..10443) /gene="ydlJL" /note="putative" /codon_start=1 /transl_table=11 /product="dehydrogenase"
	CDS	
	Query Match	7.3%; Score 83; DB 1; Length 16585;
	Best Local Similarity	49.9%; Pred. No. 5,6e-12;
	Matches 240; Conservative	0; Mismatches 235; Indels 6; Gaps 1;
OY	407 TGAGCTTCGCGGGTTGGTGTGGCAGCGCGGCTTTTGCCGATACGTCGTTACGGTG 466	
Db	10106 TGGGATTCTCTCGCTTAAGCGCGGAGCGGGGTTTTCTTGAAATACGTCCTTGCGATG 10047	
OY	467 AGGACCACATGATGTCAGAGTGCACGACTCGATGCCGAGATATGAGACTGTTGAGC 526	
Db	10046 AAGACTTTTGTTCACAACTTCTGTATGAATTTATCATATGACAGCGCGCTCGTTAGAC 9987	
OY	527 CTATTTCTGTTCCTGGCGATGCTGTGGAACGCGCTAGATTCGAGCTGGTCAGCGGCC 586	
Db	9986 CTTCTGCAAGTTGCTCTATACGCTGTCGCCCTCMAAGCAAACCTAAGACAGCGCACAAAGCGG 9927	
OY	587 TGGTCTTGGAGAGAGAGTCCCTATACGCGCTTGGCCACATCTTGCCTGTGGAAGGCCATCANG 646	
Db	9926 CTGTATTCGGCGTCGCGCCCGATCGAGACTTCTTGTTCATTGAAACGCTGAAAGCTGCCGGTG 9867	
OY	647 CGGGCAAAAATGTGTGTGTTCCGAGCCGCGCTTGATCAGAAGACAGTTTCCAAAGCAACTGG 706	
Db	9866 CAACTGATATTATACCTGTGTGAGCTTTCTCCTGGAACGCCAGCAAAAAAGCTGAGAGCTTG 9807	
OY	707 GCGCGTAGTGTTGATGCTCTTAATGTGACGACGCAAAAGCTGTCTCAAGGCTATAG 766	
Db	9806 GCGGATATATCGTTATCCGCTATAAACAGACAGATGATGTGCTG-----AGATTGAG 9753	
OY	767 TGCCGAGACAGAGGAGGATTCATGACGAGCTTCGATGCTGTGTTGTTCTCAGACATTCGA 826	
Db	9752 AACGTACAGGAGGCGGTGTGACGATTCGAAAGTACGATGCTGTCCAGAGGATGTTAC 9693	
OY	827 CCACCTCAATGTGTGCGCACGAGGAGCTTCTGCAATGCGCGCTCAATGTGGCCGCTTTGGGAG 886	
Db	9692 GACAAGCCATCCAGTCCACTACAAATTTGCCGGTGAAACCGTCATCGTCAGCAATTGGGAAA 9633	
OY	887 A 887	
Db	9632 A 9632	
RESULT 5	BSUB0004	213190 bp DNA linear BCT 26-NOV-1997
LOCUS	BSUB0004/c	Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.
DEFINITION	Bacillus subtilis complete genome (section 4 of 21): from 600701 to 813890.	
VERSION	Z99107 AL009126	
KEYWORDS	Z99107.1 GI:2632866	
SOURCE	Bacillus subtilis.	
ORGANISM	Bacillus subtilis. Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
REFERENCE	1 (bases 1 to 213190)	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertorello, M.G., Bessieres, P., Bolotin, A., Borcherdt, S., Boriss, R., Bouvier, L., Brans, A., Braun, M., Bridgell, S.C., Bron, S., Broillet, S., Bruschi, C.V., Caldwell, J.B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Cummings, N.J., Daniel, R.A., Denizot, F., Devigne, K.M., Dusterhoft, A., Ehlich, S.D., Emerson, P.T., Enlian, K.D., Erlingsson, J., Fabret, C., Ferrari, E., Foulmer, D., Fritz, C., Fujita, M., Fujita, Y., Fume, S., Galizzi, A.,

Galleron, N., Ghim, S.Y., Glaser, P., Goffeau, A., Gollightly, E.J., Grandi, G., Guiseppl, G., Guy, B.J., Haga, K., Halech, J., Harwood, C.R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Kleier-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kuita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Manuel, C., Medigue, C., Medina, N., Mellado, R.P., Munno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogilwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M., Portetelle, D., Porrolik, S., Prescott, A.M., Presseau, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B., Socolnik, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Vlati, A., Wandut, R., Wedler, E., Wedler, H., Weitzenecker, T., Winters, P., Wiput, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.F., Zumstein, E., Yoshikawa, H. and Danchin, A.

TITLE The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

JOURNAL Nature 390 (6657), 249-256 (1997)

PUBMED 98044033

9384377

REFERENCE 2 (bases 1 to 213190) Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.

AUTHORS Direct Submission

JOURNAL Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,

adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

FEATURES location/Qualifiers

1. .213190

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

33. .740

/gene="ydcf"

/function="unknown"

/note="similar to hypothetical proteins from B. subtilis"

/codon_start=1

/transl_table=1

/protein_id="CAB12375.1"

/db_xref="GI:2632869"

/db_xref="SPTREMBL:P96699"

/translation="MPQIINAIKINGVTQNCNIDVGPVHNSHRANSKYGFANFSL

GDSPTSLSLNTNIDSVDSDQDIGNPSAPISNOI"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

complement(1455. .1727)

/gene="ydcg"

gene

CDS

gene

CDS

terminator

gene

CDS

gene

terminator

CDS

gene

CDS

```

/transl_table=11
/product="Arp-binding transport protein"
/protein_id="CAB12380.1"
/db_xref="GI:2632874"
/db_xref="SWISS-PROT:P39115"
/translation="MEIIVLTINVSVEVDQTFVKHNASVVOGDIIIGIKGNACKS
TLHLIHINDAPAGQOILIKRDLKALVEDETFVAASFADQPAERKLLLEKHAVPLRDR
OJSGCEGIRLKLKQGESEADAILLDEPNHILDEPTLQIFLQOLKHNNGYVLIASHNR
VFIDEANTKRWISLDOTLLEPGFNGSGWKPERRKILQOQREYKQOKMVERIDLEADR
GLASSKKAHASTKKKGREYHRYKAKRTQDIQSKORLEKLEKAKAPLPEYTV
VRSIITLITGTRFRLQVEVLEFRLKNAPFTLOHEKVALIIGPNSSGKTTLL
NIILGGTAAGSVAWSFSAVNIIGLYLQVEFDLPLETPLELNEFTRFARGHVDQLMH
LGFPAWTFEPYRIKHSNGSGEYVRIKIMAYLLEKDVLLIDEPNLHLDPSRQLEETLS
OYSTLTLASHDRIRPLEKTKTNSVLISNNGIKQOLNDVPSRNEREELRIKLETEROE
VLGKLSFMTNDKGYEILQDAFELIKRIKELDHQDKKD"
complement(5428..7089)
/gene="ydgf"
complement(5428..5450)
/gene="ydgf"
complement(5678..5703)
/gene="ydgf"
complement(5713..7089)
/gene="ydgf"
/function="unknown"
/note="similar to amino acid ABC transporter (permease)"
/codon_start=1
/transl_table=11
/protein_id="CAB12381.1"
/db_xref="GI:2632875"
/db_xref="SWISS-PROT:P96704"
/translation="IMDDMDKMINOQTLQGRIGKRNHILQIATGCAITGTLFGSGKS
IHFGPSLREYVMTGTIGICPLKRSIGELLNGLNLYHSHFVFDQYLGDMAFITGWT
YFWPCISLTIAMADLAVLQYQWLPQWQVNPGLALILILMLNLAIVKLEGELEFNP
ALIVIALIVLALIVGLVIMKFGSGSSGPTNLMSHGGLEPNMHGELIFQWVEY
AFVGLIEVGLTAGTETPEVYFKRINNPVRYLLFVIGALLVIMSITPMIDINPSSS
PFVQVPAVAVGVGAALITNPVLTLSAASCONSVSTSTAMVSTLAKDNAPESAKLIT

```

QY	407	TGACCTTCGCGGTTTGGGTGTCGCACGCGCGGCTTTTCCGAGTACGTGCTTTACGGTG	466
Db	77631	TGGGATTCCTCGCGTTTAGCGCGCGGAGCGCGGGCTTTCTCGAATPACGTCTGTGTGATG	77572
QY	467	AGGACCACTATGGCCAAAGCTGCCAGACTCGATCTGATCCCGAGATATTGAGCACTGGTTGAGC	526
Db	77571	AAGAGCTTTTGTTCACAACTCTCTGTATGAATTATCATATAGAACAAAGGGCGCCTGTGTTGAC	77512
QY	527	CTATTTCTGTGTGCTTGCGCATGCTGTTGAAACGGCTAGATTCCAGCCTGTGTACAGCGCC	586
Db	77511	CTTCTGCAGTGTGCTCATACGCTGTGCGCTCAAGCAAACTCAAGACAGGGGACAAACGG	77452
QY	587	TGTTTCTGGAGAGGTCCTATGCGCACTGGCAACCATTCCTGCTCAAGGCCATCATG	646
Db	77451	CTGTATTCCGCTCCGCGCCGATCGGACTCTTCTTGATTGAAAGCCCTGAAGCTCCGCGTG	77392
QY	647	CGGCGAAATTTGTGTGTCGAGCGCGGCTTGATCAGAACTTTGCCAAGAGACTGG	706
Db	77391	CAACTATATTTTACGCTGTTGAGCTTTTCTCTGAACGCCACAAAGACTGAGAGCTTG	77332
QY	707	GGCTGAAAGTTCATCTCTTCAATGAGAGACGACAAATGCGTTCTCAAGGCTATTG	766
Db	77331	GGCGATCATCGTTGATCCGCTCTAAACACAGACATGTATGTGCTG-----AGATTGAC	77278
QY	767	TGCCGAGAAAGGAGATTCATGACGCGCTTGACTGCTGTGTGTTCTTCAGACATTCA	826
Db	77277	AAGTACAGGAGCGGTGTTGACGTAGACATTCAGATGACATGCTGCTGCCAGTGGTTTAC	77218
QY	827	CCACCTCAATGTGCGCAGGAGACCTTCGTGAAATGCGCGTCAATGTGGCGTGTGGGAG	886
Db	77217	GACACGCATTCACGTCCATCAATATGGCGGGAACCGTATGTCACGACTTTGGGAAA	77158

QY	887	A	887	
Db	77157	A	77157	
RESULT	6			
LOCUS	AX312470			
DEFINITION	AX312470	339 bp	DNA	linear
ACCESSION	Sequence 5455 from Patent WO0190366.			
VERSION	AX312470			
KEYWORDS	AX312470.1	GI:17897465		
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1			
JOURNAL	Leach M.D. and Shinkets, R.A.			
FEATURES	Human polynucleotides and polypeptides encoded thereby			
Source	Patent: WO 0190366-A 5455 29-NOV-2001;			
	Curagen Corporation (US)			
	Location/Qualifiers			
	1..339			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
BASE COUNT	78 a	70 c	101 g	90 t
ORIGIN				

Query Match	6.1%	Score 69.8	DB 6	Length 339
Best Local Similarity	50.4%	Pred. No. 2e-08		
Matches 170; Conservative	0	Mismatches 167		Indels 0; Gaps 0

Db	Accession	Definition	LOCUS	Result 7
QY	254	GCTCGTGATGTCACAGTGTGAACCTGTTGACAGTGCAGGTTGAACTGACCTGTCATT	CNS06K6L/c	717 bp DNA linear STS 10-JAN-2001
Db	2	GCCCTGGTGATGAAAACTTGAAGTGGGAGACAAAGTGTGTGCGACCCACAGCTACAT	LOCUS	
QY	314	GCTCGACAGATGCGGCTACACAGACACAGTGCGCCAAGACCTTGGGCTGTGTATGAGCCT	DEFINITION	T7 end of clone AT0AA002E01 of library AT0AA from strain CBS 4311
Db	62	GCAGAGACCCGATATGCTGTGGCCCCCTGTGGCCAAACGTTGACAAAGGAATGCTGCGTGT	ACCESSION	AL0402499
QY	374	GCCAGAGCGGATCTCCGACCTGCTGTGCGTGGCTGACCTTCTGCGGTTTGGTGGTGCCA	VERSION	AL0402499.1 GI:12161549
Db	122	GCAAAAGGGCTACTATTAACATTTTTCATATTTGGGGCTTGTGTGTCGGGTGTGCAGA	KEYWORDS	
QY	434	GGGGGCGTTTTCGCCAGTACGTCGTTTACGGTGGAGGACCAATGTCACAGTGCACAGAT	SOURCE	Saccharomyces servazzii.
Db	182	GGGGTGATTTGCACAAACGCTGTTGTGATGACAAATCTCAGCTGCTACAAAGTACCGGACT	ORGANISM	Saccharomyces servazzii
QY	494	CGATTCCGACCATATTGGAGCAGCTGTGAGCCATATTTCTGTTCCCTGGCATGCTGTTG	REFERENCE	1 (bases 1 to 717)
Db	242	TGCTGCCCTTACAGCTGTGGACGCTTGTGATTCAAACCGTGGCTGTGCTGCATGCAATTA	AUTHORS	Soulet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Biolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
QY	554	AACGGCTAGATTCCAGCCTGCTCAGACGCGCCCTGCT		
Db	302	GAGTCTGCGAGTTTCAAGCAGCGCTCTACGCTTTGAT		

TITLE	
JOURNAL	Malpetcu,A., Neugejlise,C., Ozler-Kalopercopoulos,O., Potier,S., Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Winczer,P. and Weissenbach,J.
MEDLINE	Genomic exploration of the hemiascomycetous yeasts : 1. A set of yeast species for molecular evolution studies
PUBMED	FEMS Lett. 487 (1), 3-12 (2000)
REFERENCE	2 (bases 1 to 717)
AUTHORS	Casaregola,S., Lepingle,A., Bon,E., Neugejlise,C., Nguyen,H., Artiguenave,F., Winczer,P. and Galliardin,C.
TITLE	Genomic exploration of the hemiascomycetous yeasts : 7.
JOURNAL	Saccharomyces servazzii
MEDLINE	FEMS Lett. 487 (1), 47-51 (2000)
PUBMED	20584717
REFERENCE	3 (bases 1 to 717)
AUTHORS	Genoscope.
TITLE	Direct Submision
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia soditrophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES	location/Qualifiers 1..717 /organism="Saccharomyces servazzii" /strain="CBS 4311" /db_xref="taxon:27293" /clone="AT0AA002E01" /clone_1lb="AT0AA" /note="end : 77"
misc_feature	complement(<10..>333) /note="similar to Saccharomyces cerevisiae ORF YAL060w [FUN8] ; similarity to alcohol/sorbitol dehydrogenase } similar to Saccharomyces cerevisiae ORF YAL061w [FUN50 similarity to alcohol/sorbitol dehydrogenase] "
BASE COUNT	222 a 146 c 126 g 211 t 12 others
ORIGIN	
Query Match	5.9% Score 67.4; DB 11; Length 717;
Best Local Similarity	53.1%; Pred. No. le-07;
Matches 171; Conservative 0; Mismatches 142; Indels 9; Gaps 1;	
OY	1 ATGAAGCTTACTTTATTAGCGTACCAACGATATTCGTAACGGTCCCTGAA 60
Dd	333 ATGAGAGGTGTGCTATTTTAAAGAGGTGATATACATTTCACAGATAGTTGCCAGAA 274
OY	61 CCGGATCACAAGAACCCACAGCATGCACAAAGATCAAGTCAGTATGTGAGATCTGGCC 120
Dd	273 CCMAAGCTTACTCCCTCGATAGTCTACTTATAAGCCCCCTCGTTGTAGGATGTGTGT 214
OY	121 ACGGACTTGAAAGAATTCACATATTTCTGGAGGCTCGTTTTTTTCCCTAACAAGCACACC 180
Dd	213 AGTGACTTGCAATGAG-----GCAAGCCATCCAAATCTTTTCCCTAATAATGGCCAC 163
OY	181 AAGGACAAGATTTCCGGATACGAACCTCTCTCTGCTCTGACATGATATTTAGCGGAACG 240
Dd	162 AGACATTAAGCTTAAGTGAANAATGAACCTTCCCARGCTCTTGACACAGAGTAGTGATAT 103
OY	241 GTGGTCGAGGTTGGCTGTGCTGACAAAGTGGAACCGGATGAGATGCGACAGTTGAA 300
Dd	102 GTGAAGCTATTGGACACCGCTGACTCAATTTAAAGTAGAGGAGACCGAATCTGTCGCAA 43

[illegible]

OY 417 CGTTTGGTGTGTCAGACGCGGTTTGGCAGTACGTGCTTACGGTGAAGACACAT 476
 DB 439 TGCACCTCCACCGGTTTACGTGCTTACGTAACCAAGTGTACACCTGGGACATATG 498
 OY 477 GGTCAAGTGGCAGACACTGATGATCCGAGATATGGAGCACTGGTGAAGCTATTTCTGT 536
 DB 499 CTTCAAAATGCTGAGATGATGATTTGGAGAGAACATGATGTAACCACTTAGTGT 558
 OY 537 TGCCTGGCATGCTGTGTAAGCGCTAGATTTCCAGCTGGTGGTGAAGCGGCTTCTTGG 596
 DB 559 TGGTGTCCATGCTGTGCGCCAGCTGAGGTGTCTGTAACCAACGTTTGTGTAATGGG 618
 OY 597 AGAGATGCTATGCGCCCTTGCACCTTCTGCTGCAAGCCATCATCGCGGCAAAAT 656
 DB 619 AGCTGACACTATTTGGCCCTTGTACAAATGTGGCTGCTGCGGCTTTCAGTGTGCTAGAT 678
 OY 657 TGTGTGTTCCAGCGGCTTGTATCAGAACAGACTTGGCAAGAACTGGCGCTGAAGT 716
 DB 679 TGTATTTGTGATGTTGTATGTAAGACCGTTTACCGGTAGCAAAAGCACTGGCGGAGATGA 738
 OY 717 GTTGCATCTTCTACATGTGACGACGCAATGCTTCTCAAGCTATGTTGCCGGAGAA 776
 DB 739 GATTGTACAAGTGAACAACTTAGAGATGTTGTTCTGAGGTGAACAGATTCAGAA 798
 OY 777 CG-----AGGATTCATGCGACGCTTGCAGCTGCTGCTGCTCAGACATTTCAC 827
 DB 799 AGCTATGGGCTCAACATCGATGTAACATTGCACTGTCGGGCTTTTAAACAAACCATGTC 858
 OY 828 CACCTCAATTTGTGCGCACGCGGACCTTGTGAATGCGCGCTCAATGTGG 874
 DB 859 GACAGCATTTAGCACGCACTGCTGTGGGGTAAAGTGTGCTTGTGCG 905

RESULT 10
 LOCUS AY085213 1347 bp mRNA linear PLN 25-JUN-2002
 DEFINITION Arabidopsis thaliana clone 13875 mRNA, complete sequence.
 ACCESSION AY085213
 VERSION AY085213.1 GI:21403923
 KEYWORDS FLI-CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 REFERENCE Eukaryotes: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1347)
 Haas,B.J., Volfovsky,N., Town,C.D., Troupkan,M., Alexandrov,N.,
 Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
 TITLE Full-length messenger RNA sequences greatly improve genome
 annotation
 JOURNAL Genome Biol. (2002) In press
 REFERENCE 2 (bases 1 to 1347)
 AUTHORS Brover,V., Troupkan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
 Feldmann,K.
 TITLE Full-length cDNA from Arabidopsis thaliana
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1347)
 AUTHORS Brover,V., Troupkan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
 Feldmann,K.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
 Malibu, CA 90265, USA

This clone sequence is one of 5,000 Ceres full-length cDNAs made
 available to TIGR and Genbank. The following quality assessment of
 this set was done by comparison with known proteins: two percent of
 the clones are estimated to be 5'-truncated; less than one percent
 are 3'-truncated; approximately two percent represent alternative
 splice variants, including unspliced introns and spliced exons; one
 percent may contain premature stop codons; five percent may have
 frame shifts in a coding region. A sequence is considered to be
 5'-truncated if it lacks the translation initiation start (ATG). A
 sequence is considered to be 3'-truncated if it lacks the

C-terminal end of the encoded protein. Please note that these cDNA
 sequences are derived from the Ws or later ecotypes and therefore
 may contain polymorphisms when compared to sequences from Col-0.
 Genet carried out the library production and sequencing of the
 full-length clones. Ceres, Inc, carried out the clustering of the
 5' sequences, selection of clones, and sequence assembly.

FEATURES

1. 1347
Location/Qualifiers

/organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="13875"
 136..1230
 /codon_start=1
 /product="sorbitol dehydrogenase-like protein"
 /protein_id="AA02446.1"
 /db_xref="GI:2153353"
 /translation="MGKSGSOGSGSKEVEENNAALVGIINTLKIQPLLPSPGPHY
 RYMKAVGICGSDVHYLTKMRCAEFVYKEPMVYIGHECAGIIEVEGKHLVWDRVA
 LEPIGSCMRNLCEGRVYNLCEPKFPATPPVHSLANOVHPADLFKLPENSLSE
 GAGEPIPLSVHACRAEVGPEPTVLMVAGPGLIYMLAAQAFSVPRIVYVDENR
 LAYAKOLGADETIYQYTNLEEDGSEVQIQKAMSNTDYPEDCAFRTKSTALAAIR
 CGGRVCLVGHGIMTYPLTPAAREVDVYGVFRITKNTWPLCLEFLTSGKIDVPLIT
 HRFESQKEVEDAEFTSARGSMAIKVFNL"

CDS

BASE COUNT 358 a 235 c 354 g 400 t
 ORIGIN
 Query Match 5.9% Score 67; DB 8; Length 1347;
 Best Local Similarity 47.18; Pred. No. 1,4e-07;
 Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

OY 357 TGGGCTGTATGCGCTCCAGACGCGATCTCCGACATGCTGCGTGCAGCTGATCTTG 416
 DB 462 TTGGAGATGCAATCTCTGACGAGGAAGACGATACCACTATGTCCAGAAATGAATCTT 521
 OY 417 CGTTTGGGTGTCACGCGCGGCTTCCGAGTACGCTGTTAAGGTGAGACACAT 476
 DB 522 TGCAACTCCACCGGTTTCACTTGTAGTAACGAGGTGTTCAVCTGGGATCTATG 581
 OY 477 GGTCAAGCTGCACAGCTGATTTCCGAGATATGGAGCACTGTTGAGCCATTTCTGT 536
 DB 582 CTTCAAAATGCTGAGAAATGATGTTGGAGAGAGCAATGTGTGAACCACTTAGTGT 641
 OY 537 TGCCTGGCATGCTGTTGAACGCGCTGAGATTCACGCTGTCAGAGCGGCTGTCTTGG 596
 DB 642 TGTGTCCATGCTTGTCCGACGAGTGTGCTCTGAACAAACGTTTGTGTAATGG 701
 OY 597 AGAAGTCTATGCGCTTGCACCACTTGTCTCTGCAAGGCCATTCATGCGGCAAAAT 656
 DB 702 AGCTGACCTATGTCCTTGTACAAATGTGGCTCTCAGGCTTCACTGTGCCATGAAT 761
 OY 657 TGTGTGTTCCAGCGGCTTGTATCAGAACAGCTTTCGCAAGAAAGAACTGGGCGTGAAGT 716
 DB 762 TGTATTTGTGATGTTGTGAGAAACGTTTGGCCGTACCAAAACGTCGGCGCAGATGA 821
 OY 717 GTTGCATCTTCTACATGTGACGACGCAAAATGCTTGTCAAGGCTATGTCGCGAGAA 776
 DB 822 GATTGTACAAGTGAACAAACTTAGAGAGATGTTGTTCTGAGAGTTGAACAGATTCAAGA 881
 OY 777 CG-----AGGATTCATGCGACCTTGCAGCTGCTGTGTTCTCAGACATTCAC 827
 DB 882 AGCTATGGGCTCAACATCGATGTAACATTCAGTGTGCGGCTTTTAAACAAACCATGTC 941
 OY 828 CACCTCAATTTGTGCGCACGCGGACCTTGTGAATGCGCGCTCAATGTGG 874
 DB 942 GACAGCATTTAGCACGCACTGTTGTGGGGTAAAGTGTGCTTGTGCG 988

RESULT 11

LOCUS AB025969 1457 bp mRNA linear PLN 04-APR-2000
 DEFINITION Prunus persica mRNA for NAD-dependent sorbitol dehydrogenase,
 complete cds.
 ACCESSION AB025969

QY. 770 CGGAGACGAGGATTCCATGCAGCCCTTCGACTGCTGTGTTTCCTCAGACATTACCA 829

```
complement(1934, .2950)
/gene="AGR_L_266"
```

```
/note="(AE005211) putative permease component of transport
system, probably ribose specific"
/codon_start=1
/transl_table=11
/product="AGR_L_266p"
/protein_id="AK88704.1"
/db_xref="GI:15158438"
/translation="MRRLIGHTEFTLLVVMVLLCGSPATDRFLTISNAPDLNV
SAVNIIFAVGLVLLVLSGIDISFAVAAYOVYVLAALNAGGNAEGRTIANGANG
LSGFGVAPLWRLNIISIVATITFTIFPGLMEFTKGVSTIDLPMLSTRTVETER
EMPDGSEVELTLPVYVNIICCVATWMTSRTYGRKLAFGDNEGARGINGAMH
YISGMLGVLNAGLNGQAHQAEVNPALYVLAITVLEGARLGGGKGVICG
VLCVLAWSITONGNLIMGVSPFAKMTVGAIIIVAILTSTREDKLPLRTSAKKG
SAOR"
complement(2943..4487)
/gene="AGR_L_268"
complement(2943..4487)
/gene="AGR_L_268"
/note="sugar ABC transporter, ATP-binding protein -
Thermotoga maritima (strain MSB8)"
/codon_start=1
/transl_table=11
/product="AGR_L_268p"
/protein_id="AK88705.1"
/db_xref="GI:15158439"
/translation="MPSATSAMLSRRKERYAINSAPLELSKNVKTFFGGVRLKGV
FEVPGFVHCLAGENGCGKSTVIKIVGVKPPSDAELTFDGPDIAMPTPLNOSGI
QVIMODALFDEMTVAENIGFOYAVNGRGLVPRKAKNAEKALATGSPDLRPL
KELPIAOROVATARALVGARLVFNDERTASLTOSTDYLLIDIVRLSASGAVVV
SHLAEVLETSIDRTVLRDGLVGFVEEGMTGRNFPDSAVIAAODHDKP
VLSVRLTRAGEFEDVSFLRGETLGTIGLACGRTLELATLFGMHRPOSEVLID
GKVVDFHSNDAIRVGVAYLSEDRSLGLOPOSIDNLVNASLDKLSSGLSTSK
ADVSRWISALGVKIGLPEDPRTLSGNGOQVAIAKMLIGPKIILIDAPTVGVG
ARAGIFEIVKLAEGSLIVISDEAPEVYFNDRVIMHEGRFHAITYDRRLSTEL
EALIVA"
complement(4551..5528)
/gene="AGR_L_271"
complement(4551..5528)
/gene="AGR_L_271"
/note="(AE005211) putative periplasmic binding protein,
probable substrate ribose"
/codon_start=1
/transl_table=11
/product="AGR_L_271p"
/protein_id="AK88706.1"
/db_xref="GI:15158440"
/translation="MKKIITAAVGLSLALISSAARFEGPVGVVVKIGTIPFNMAEA
GIREQSNKLGVDGFVNGPTSDALVLRALIEDLIQKVPETIGVVPNDAKVLEVLKA
QAAGIVITIHESPELGADMDPELAGAKGEGHAKLLEKMGKGSYAVFVGLTVP
LHNAWADAIAIYIKANYPDKLVGDYVAEDLDKSRSTALDLMSANADLKGLAFGS
QCGIGAGRAVEERRKDKGVFVIGFSPGGGAKLIKSGALTGCFMNMKQAGVFITLA
DRAKETPRAGGTIEGLTIDPKNTIYVDOLKIDKESIDKLSMGL"
5768..6847
/gene="AGR_L_273"
5768..6847
/gene="AGR_L_273"
/note="fructose transport system repressor FRUR PA3563
[Imported] - Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_273p"
/protein_id="AK88707.1"
/db_xref="GI:15158441"
/translation="MTGISSKKATIDYLSILSGASATYSAVINGSMRKRRISEETAD
KILSLAKORYTNIQARGLRSKGLVLLPVYNRFPSSMAOTFECQAKRRLGSL
MYVSGRDRPEERERTYETILATISIDLFTAGVTDPGVQVOCARALPVAINDLPKPF
ASSVLSNNRGAELITLALILHAAKGSIGPDVILFGHDDHSAERDILFAGAAAD
YFGEVGGDDIEITGYSPHMTMAFERFRGRGLPCFVNINSINEGGLRFEGRADG
EAFGLIVGCFDPDPAFSLPPVYMIKPDIDIAQMLEKGFELLEENRETEVETIIEQL
IPRPLALEGPDIDMPVALRRMAK"
6878..7267
/gene="AGR_L_275"
6878..7267
```

```
/gene="AGR_L_275"
/note="hypothetical protein PA0015 [Imported] -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_275p"
/protein_id="AK88708.1"
/db_xref="GI:15158442"
/translation="MTSTALDVPIICNRHHNGTITIEKLMVLESGRDSPVLRSL
AKALMDROPEAOHLOEAIIRONPDYSAAWALAGECERELGDHEGAAMAGIVAVA
IRGDDIQAKRMEAFRRRLQTRLSAS"
7391..9274
/gene="AGR_L_278"
7391..9274
/gene="AGR_L_278"
/note="ACETOIN CATABOLISM REGULATORY PROTEIN"
/codon_start=1
/transl_table=11
/product="AGR_L_278p"
/protein_id="AK88709.1"
/db_xref="GI:15158443"
/translation="MGFSDHIKEIETVGGGRNTGRNAVLESNRCLDTYQDPTKAR
EAVIVSETRLREHQOAEHLIARSGLERTYRQVAEQNVLLSDPGQVTEFLGP
SFNNLRKAGLYGSEWSEPRAGCAVGIATGSELTTHQDTPVTRPSTCAP
IYDPOGLATVILDISLSPLTKTSOMAHIVSYRRIELANLMASSRHDIYLRPA
GAEPLDPEALSDVSGSRITGTHAARLILASRGDMRPERLGRVVEFEA
GDELSLITRVSILPQDRIVYRGTVLYAALBPQGRSSVPILARSPOPVTERAIGD
GPVRNLRKRVKLESPARLPILQGGTGTEKELIARIHDSVSGFVFNCAIPE
OLIESELEFGLPGAFGLAKGRKGLVEEAGGTLEFIDEIGDMPAQSRLVLEAG
EVLPGVSGTPQKDFRVYISASHPRLDVTGAFREPLYRANAAYLTUTPALSERDF
PMLDLRLEKHGMRPRLTISNARLTLLIHHRPGNIRELDNAIAPAAACDGLTIV
EDLPOLTRASSITPMDSKOTDRLALLACNGNISSEARLVGDRITLHRMRRLIG
KPI"
9311..10504
/gene="AGR_L_281"
9311..10504
/gene="AGR_L_281"
/note="(AE006323) 2,3-butanediol dehydrogenase (EC
1.1.1.4)"
/codon_start=1
/transl_table=11
/product="AGR_L_281p"
/protein_id="AK88710.1"
/db_xref="GI:15158444"
/translation="MAGKRTGDPQNPATISGLMRKPFMWHGHTFSGIKCHRGEMRAL
RFLAANDLIEDIAEPKRPQGVVLRNRFVIGCTDLHEYSYPIFETPEPFTGA
HGPOILGHEFGVVEAIGDVTSVNGDIQPLIMPBGDVFADRGLEHISTOLAL
VGLSMDGGMAEALVNEYNVORIPDEMTDEBALVPEVSAVYACDRGGVAGNSVL
VIGASPTIGMTLLAARAAGATOFVSDLDNARLEIARVYTRQVITINPRDNYGDVVR
SMEKGVGCDVAIECGENHAKACVDAYRKQGVVYOTGLHPHENTIDMFVYTRDLE
IKSMAYPTHYMVRVIRLIASGLLPATKIVTKVITLTDTAVTEGFDALDPAGTHAKIL
IDLSK"
BASE COUNT 2210 a 3246 c 3137 g 1989 t
ORIGIN
Query Match 5 68; Score 64.4; DB 1; Length 10582;
Best Local Similarity 47.58; Pired. No. 1e-06;
Matches 225; Conservative 0; Mismatches 246; Indels 3; Gaps 1;
```

Db 9952 TAATAGCGTTTGGTCACGGGTGCGGCCGATGCGCATGTTGACGCTTTCGCGCCCG 10011

636 AGGCATTCATGCGGGGCAAAATGTGTGTCCGAGCGGCTTGATCAGAGACAGTTTGC 695

10012 CGCGCGGGCGGTACACCACTTTTGTCTCCGATCTGAGATGACGCCCGCTGAGCTTGC 10071

696 ---AAGGAACTGGGCGCTGAAAGTGTTCATCTTTCATATGTGACGACGCAATGCTCT 752

10072 CCGCATGTCATCCGATGATGATCATCACATCAACCCCAAGCGGATPAATGCGTGATGT 10131

753 TCTCAAGCATGATGTCGCGAGAGACGAGGATTCATGACGCTTGACTGCTGTGT 812

10132 CGTCCCTGCGGACCGAGGAGGAGAGTCCGCTCGGACGCGGATGATGATGATGAT 10191

813 TCTCAGACATTCACCACTCAATGTGCGCACGAGGACCTTTCGATGCGCGT 866

10192 TGACATGCGCTCAGGCGCTGCGTGCAGCGCGGAGGAGGCGGCGTGTGT 10245

RESULT 13

AE009402/c 11742 bp DNA linear BCT 20-DEC-2001

LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 172

DEFINITION of 187 of the complete sequence.

ACCESSION AE009402 AE008689

VERSION AE009402.1 GI:17743239

KEYWORDS

SOURCE Agrobacterium tumefaciens str. C58 (U. Washington).

ORGANISM Agrobacterium tumefaciens str. C58 (U. Washington)

REFERENCE 1 (bases 1 to 11742)

AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmeri, A., Raymond, C., Rouse, G., Saenphimachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.

TITLE The genome of the natural genetic engineer Agrobacterium tumefaciens C58

JOURNAL Science 294 (5550), 2317-2323 (2001)

PUBMED 11743193

REFERENCE 2 (bases 1 to 11742)

AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmeri, A., Raymond, C., Rouse, G., Saenphimachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.

TITLE Direct Submission

JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

FEATURES

source location/Qualifiers

1. 11742

/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"

/strain="C58"

/db_xref="taxon:180835"

213. 3002

/gene="acnB"

/note="Atu4734"

213. 3002

/gene="acnB"

CDS

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="aconitate hydratase 2"

/protein_id="AA145528.1"

/db_xref="GI:17743240"

/translation="MNLYIDVLAIEKSEPTGLAPKPIDGALFAEIIIEIDTSEY RADLKPFITYITLTGTTSAAGVKAAPLKQIILGGLVLPETSPALFLLSMKGGPSI KVLIDVAGNDAAIDAKAGEVLTQVTEIAAGDILSTDLISGNOAHSRDLQIGQCMISPEADAEIV FETKLPDVDDIEIKIVTEIAAGDILSTDLISGNOAHSRDLQIGQCMISPEADAEIV ALQOQHPDKRVMIAEKGTMGVSSRMSGVNNAVLWQKASPVPEVNAIVAGIN GISPFPATVTVTGIGIINKWYKLGSEDKPILINDGNPILIOKYSVEGTGLKD AKNRRLDRENGTELVDAESPTPOKMEFMRGSSVAIVFGKLOTFAAQITGVAATGV FAPNKEISIEGQGLAVKIFIRNNAVGTGPKVHAGSDVAVKNIYGSODTGLMTA OELDMATVTSPLVDGAYOSGCHRSASVWDKKAQANTPKLSEFMHNFVGIVRDPKV IYHMTVDVIAKMTDITVDRAIITIGGDSHTRMSGVAFAGDSGTVALATANGEMTI POSVAVTEFGTMQPMDFRDVVAHQAOMLQHDNDVFOGRIIEVHIGTLADDAFTF IDTVAEMKAKASISISEDETLIESLEIAKRSIQIMIDGMDNAQTGLIKRADQRI AEIRSGETPALKPDANAKYFAEVVVDIINEPMIADVDVNNNDVSRVYTHDITRPS YGATKKVDLGEFGSCWYHKGDKVIAQMLNMEKTEGKVEKAPLVAAPTNIIIE LKAEGDWEIILQKYSGEFDDVNPNTSEYDNLITLERPCNLCMGNOEAAAGDVF LATSTRLEQGRVVEDSAKKESSLASIPVYVLSAILGRPSIDIEYSVAVSGIDLTF APKRTPIDSQSVH"

complement(2999..3196)

/gene="Atu4735"

complement(2999..3196)

/gene="Atu4735"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AA145530.1"

/db_xref="GI:17743241"

/translation="MKCYPTVASADNSARESPNLIHQIIRNHLHTNMAEIDRCV HSSDGQGRIRHPRKAPDA"

3399..5354

/gene="Atu4736"

3399..5354

/gene="Atu4736"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="methyl-accepting chemotaxis protein"

/protein_id="AA145530.1"

/db_xref="GI:17743242"

/translation="MRPMSFLKNTKIKTKVFEVLSIMSLSGISVYSLOYKTTDA VYSDPIGHEALAAVYNARTSGNLNLMQIIRASINDPASSDPEFAVYTFPRAROLE ERQNKIMELVPAKTDARDILKGVYEIEIGNQYITLMQAKQAQOAMALNVLRKIT EVSPKISAGNEOLIVMDGRRLTASTNTITWGLITLALVSTALILGLVSRGI TTPIRLARARMSLSLAGDTASEIDMDRDEVGQAQAVQARENAIKARVLENTGA NMSMEKRIERERKQAEADVDFVANNLITALSRIEGVTAIADPFPVNLGIR GDFNRAEOLOSTIFQVANOAGIDAGNNLITARSADDLARTEOQAARETAALAE ITTYWQDSTKRAQEGHIVGRAKIGAEKGEVQAVSAMEDIARSNAKETKILITNSGO ETAFQTNLILAAQVEARAGAGGAPVAVQVEELAORSNAKETKILITNSGO VOQGVQVGEGERALVELVTEQELNRRVAAIASEAQOSSLOOINAVNOMODTPA KNAAVESTASHLARPASSLNRLIQFLKSEGVETAAPMTASAPPAVSPA RALGGRIRAEFSGNALNTAADNMEEF"

5553..5735

/gene="Atu4737"

5553..5735

/gene="Atu4737"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AA145531.1"

/db_xref="GI:17743243"

/translation="MIAVYIDFIMFCAGLMTLVFGYLALSKNPAYQPRVYHFKM MCPMLIGAIILAIAS"

5940..6569

/gene="cat"

gene

CDS

Query Match	5.6%	Score 64.4	DB 1	Length 11742
Best Local Similarity	47.5%	Pred. NO. 1e-06		
Matches 225	Conservative 0	Mismatches 246	Indels 3	Gaps 1
Qy	396	CTGTGCGTCGTGAGACCTCTCTCGCGTGTGGTGTGTGTCACAGCGCGGCTTTGCGCGATACGT	455	
Db	9270	CAGCAGCGCACTGCGACACTGGTGTGGGCTCTCAGCTGAGGAGCGCGGCGATGCGCGAAGCCGC	9211	
Qy	456	CGTTTACGGTGAAGACACATATGTACAGCTGCCAGACTTCGATTCGCCAGCATATTGGAGC	515	
Db	9210	ACTGTTCACAGATACAAAGCTCCAGAAAATTCGCGACGATGACCGCAGAAAGAGGCCGC	9151	
Qy	516	ACTGGTTGAGCCTTATTTCTGTGCTGCTGGCATGCTTTGATGACGCGGTACATTCCAGCCGCG	575	
Db	9150	ACTGGTGGAAACCCCTGGCGCGTTCGCCGTTTATGCCGTATGTGCGGTGTACCCGCGG	9091	
Qy	576	TCAGACGGCCCTGGTTCCTGGAGAGGTCATGCGCCCTTGCCACACATTCCTGTCTGCA	635	
Db	9090	TATATAGCGTTGTGGTCACGGGGTCCGCCCATCGGATGTGAGCGCTTCTGGCGGCCG	9031	
Qy	636	AGCGCATCATGGGGCAAAATTTGCTGTTCGAGCGCGGCTTGAATCAGAACAGTTTGC	695	
Db	9030	CGCCCGCGGCGCTACCCACGCTTTTGTCTCGATCTGAATAGCCGCCGCTCGAGCTTGC	8971	
Qy	696	---AAAGCACTGGGCGCTGAAAGTTCGATCTTCTCATATGACGACGCAATGCGCT	752	
Db	8970	CCGCAATGTCATCCGTATGTCATCACCATCAACCCCAAGGGCGATATGTGCGGTGATGT	8911	
Qy	753	TCTCAAGGCTATAGTGGCCGAGAACGAGGATTCATGACGCTTCGACTGCTGTGTGT	812	
Db	8910	CGTCCGCTGGCAACCGAAGCGAAGTTCGGCTGGACGTGGCATTCAAATTCGTGGGTAA	8851	
Qy	813	TCTTCAGACATTCACCACTCAATTTGTGCGCAGCGGACCTTCTGGAATTCGCCGT	866	
Db	8850	TGAACATCGCGCTCAAGGCGCTGCTGTGACCGCGCTCGCAAGAGGGCGTCTGCTGT	8797	
RESULT 14				
AB015478				
LOCUS	AB015478	81347 bp	DNA linear	PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MSG15.			
ACCESSION	AB015478	BA000015		
VERSION	AB015478.1	GI:3241926		
KEYWORDS				
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitau1 P1 clone:MSG15.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryote: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta; Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eucosids II: Brassicales; Brassicaceae; Arabidopsis. 1 (sites)			
AUTHORS	Nakamura,Y., Seto,S., Asamizu,E., Kaneko,T., Kotani,H., Miyajima,N. and Tabata,S.			
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. VII. Sequence features of the regions of 1,013,767 bp covered by sixteen physically assigned P1 and TAC clones			
JOURNAL	DNA Res.	5 (5),	297-308	(1998)
MEDLINE	99087489			
REFERENCE	2 (bases 1 to 81347)			
AUTHORS	Nakamura,Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (17-JUN-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakam@kazusa.or.jp, tel:81-438-52-3935, fax:81-438-52-3934)			
COMMENT	Address for correspondence: kaos@kazusa.or.jp for the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/c91-bin/sgd-graph.cgi?c=MSG15 'product' or 'note' qualifiers. Genes that have no significant			

protein similarity are described as 'unknown protein', the software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/grail-1.3/>), GENSCAN (Chris Burge, MIT, <http://CCR-081.mlt.edu/GENSCAN.html>), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and SplicePredictor (Volker Brendel, Stanford University, <http://gremmlin.zool.jastat.ehu/cgi-bin/sp.cgi>). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MIM18 and the 3' clone is F1P19.

FEATURES

Source

exon

CDS

Location/Qualifiers
1..81347
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MSG15"
/clone_lib="Mitsui P1"
complement(1..402)
/note="CDS is reported in Acc# AB025623
gene_id:MIM18.7
P111705129
strong similarity to unknown protein"
/number=1
/evidence=not_experimental
join(2611..3415,4019..4413,4496..5056)
/note="gene_id:MSG15.1"

CDS

/codon_start=1
/evidence=not_experimental
/product="mandelonitrile lyase-like protein"
/protein_id="BAB11041.1"
/db_xref="GI:10177728"
/translation="MTITSCITIHIFRNILFALIFYGSCYCDKGTILISYKNLILV
YISIFPELVKLETTFMKDKATLAPKNSPYIIIGGTAGCALATLSOAVSLVER
GSPYENPRTADNGSVNTILNTPMSMQLFTISGQVYTRRVRVGGSGVINGFYS
DRIANDYEEEMEEVEAAYEVEKIVREPOVYIMOKAFKGLIAGSPNGSTY
DALYGRIGGTIFDRAGHRTANLLEYAPNRIYVYLHSAVKVLETTBEKKAIEV
LFEDANGVFEKANKATNEVILSAGALSQPLMLSGVPAVLEAHGVNPLVDQ
PAYGOGADPNMVAIPSPQVELSLIQAVGITKFDPSYIEGLSGLSFEDITREFD
GVNLINETSHTSRKILDSIAVLKSPVITIKVSNKGLIFOKYDGPASKHMLLR
NTNPDPNSVTENVYOEEDLNCKVGKNTITIMINSKASKYKYGVARBELTMLL
ALPINLRPHVTSARINIKQRCIDTIVTSVMHRYHGGQYKVDKNYKVLGIDGLRVLDG
STFKSPGINPQATVMMLGR
join(6167..6204,6527..6560,6653..6825,6914..6965,
7061..7128,7215..7284)
/note="gene_id:MSG15.2"
/codon_start=1
/evidence=not_experimental
/product="DNA-directed RNA polymerase II subunit-like
protein"
/protein_id="BAB11042.1"
/db_xref="GI:10177729"

CDS

/translation="MADEDYNDVDDIGEDEPAPEETEEGVEDVEKENDVNGEPI
EAEDKVTPEVQPRKTSKEMTKERKRIITGRALQISNAPYVVELEGTDULEIAM
KELQRKIPFTIRYRLDPSFEENGDELLIVDSWKQVGD"
complement(join(7728..8294,8400..8800,9212..9853,
9966..10047))
/note="gene_id:MSG15.3"
/codon_start=1
/evidence=not_experimental
/product="mandelonitrile lyase-like protein"
/protein_id="BAB11043.1"
/db_xref="GI:10177730"
/translation="MGLOILMLFKRIITYIFHASCSPDKAGYSPMKDATLAPMFA
RQYITIGGSGCALATLSOAVSLVERGAPYNPATIDENATLISNTPS
WSOLFISEDEYVTRAVVGGSLNAGFTTRAGDEYVTEKTEKTDVEAAYEMVEK
VARQPVAGMOTAKDGLERAGEPVYNGFYDHYTGKGTIFDRAGHRTADLLE
YANPNTIVYVLAHSVRLFTTQGRPRKAYGVIFQDANGVLAHKAELKSNMEVILS

CDS

AGATGSPOLLMLSGIGPAHLAANGIKPLVDHPVGGQMGNDPMNAIFTPSPPEVY
SLIOVGTTFKEFSYIEGASGYISVSWTRREFPDVLTNINLSTSTSTQSTIT
DEPNINPLNATTRAGILCKINGISRGHLELRNPNDSNPSRNYVQEPDQI
CVEGINTITIKYINSAKSFERYPATHTIGLIDMLSVPTNLRPHITSMFNLROFCD
TWTMTWHGGQYGRVYDKNRYVLGIDSLVDIGSTFLKSPGINPQATVMMLGR"
join(12204..12401,12491..12604)
/note="gene_id:MSG15.4
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11044.1"
/db_xref="GI:10177731"

CDS

/translation="MSGCVLAARYRDLAAVYKHGKEDKHSHFLDPVQOEPRKA
NSEKINLARNTYILNLSIHSKDLIFSNIYAVDRTEERKRVLNKSAASVGLRPEVYE
P"
join(13308..13425,13521..13588,13687..14214,14299..14486,
14580..14772)
/note="gene_id:MSG15.5"
/codon_start=1
/evidence=not_experimental
/product="sorbitol dehydrogenase-like protein"
/protein_id="BAB11045.1"
/db_xref="GI:10177732"

CDS

/translation="MGKGMSSGSEBSKYEEENMANIYINTIKIOPILPSPGHVY
VTRAVAGIGSDVHYLTKRCADVVEAPVYIGHECGITIEEGEYKHLVYDGRVA
LEPGISVCNRCNCRGRYNLCPEMKEFATPEVHGSLLANQVHPADLCKLEPNVSLER
GAMCEPLSGVHACRAEYGETNVLVAGAGTIGLVTMLAARASVPRIYVDENR
IAVAROLGDEITVOYVTNLEDVGSVEVOIQKMGNSINDVTEFCAGFNSTALADR
CGKVCIVGMGHGIVTPLTPPAAREVVOVGFRKNTMPLCEFLTISGKIDVPLIT
HFGFSQKVEDEAFETSARGSNATKVMNL"
complement(join(15138..15282,15363..15460,15557..15616,
15695..15814,15916..16062,16155..16256,16376..16465,
16839..17390))
/note="contains similarity to myosin heavy chain kinase
gene_id:MSG15.6"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB11046.1"
/db_xref="GI:10177733"

CDS

/translation="MDFDLNGKNKRVFNRLGGGSTRPMAFTDRORVCFHRAGRC
NSPCPYLRELPFGPGGQGYTKNKVAEESGFAGSHRGGFNGNSSWGRF
GNRTVYTKERKYNVQDNCYQKCRHLKMSGSEALITODGHEKLVSGIALP
SSDKIYTSKQETRLRWMDCAAGCTGCTGKGLGELVSEBPMILVQMPNLVKYMI
ENADDSLSGPVQGYSLVYGDPLFAGTODSILAMRNATNPFESASITGTLA
VYTLVYAGARKLSSGMDKTIKWSLDNLQCIQTLVHSSVNSLTCMDFLISGLDN
VILIMAIIEGLGMEVYTIHKEHGVIALCGVDAKAPVLLCACNDNLTLYDPSFT
EKGKIPAKOEIRAIQIGPGIIFETGDTGQVYVWVWCTEPTALP"
complement(18601..19275)
/note="gene_id:MSG15.7"
/codon_start=1
/evidence=not_experimental
/product="Ap2 domain transcription factor-like protein"
/protein_id="BAB11047.1"
/db_xref="GI:10177734"

CDS

/translation="MNPYSTPDSFLSTSDHRSVDSSECPKIASCPKRRAGRK
KRETRHPITRGVRONSGRKWCVEAREPKKSRINLGTPEYEMARADVAALRGG
RSACLNFAASAMRLRIPETTCPEIOLKASSEAMAFONITIEGSKTAAEAAEABG
VERGEERAEQNGVYFMDDEALGMPNFFEMAGSMILLPPEVGMNHADFQGDVDS
LMSFDE"
complement(join(20512..20836,20925..21052,21150..21278,
21377..21455,21539..21702,21778..21954,22033..22155,
22230..22312,22406..22505,22550..22607))
/note="gene_id:MSG15.8"
/codon_start=1
/evidence=not_experimental
/product="importin alpha subunit"
/protein_id="BAB11048.1"
/db_xref="GI:10177735"

CDS	WISNITAGTREQIOSVIDANLIPILVNLADQIDFYMKKEAVNAISMALNSHQIK YMAEQSCIKQICLILVYSDERTITKCLDLENNKAGEKSEEDNPNYCLIEDAE GLEKISLKNKNDIDYERAKYKILVTWFEEDDENNNNNRCDDVDQV"	complement(23114..24304) /note="db11BA78737.1 gene_id:MSG15.9 strong similarity to unknown protein"
CDS	complement(25680..26378) /note="contains similarity to Ap2 domain transcription factor gene_id:MSG15.10"	
Query Match	5.68; Score 64.2; DB 8; Length 81347; Best Local Similarity 48.5%; Pred. No. 1.4e-06; Matches 177; Conservative 0; Mismatches 188; Indels 0; Gaps 0;	
OY	357 TGGGCTGTATGGCGCCGACGAGCGATGCCAAGTGGCTGCTGCGTGGAGCTTCTG 416	
Db	13827 TTGGAGATGCATCTTGCAGGAGAGACGATACACCTTTGTCAGAAAGATCTT 13886	
OY	417 CGGTTGGGTGTCGCCGCGCGGCTTTGCCGAGTACGCTTACGCTGAGACCAT 476	
Db	13887 TGCMACTCACCGGTCATGGCTCTTAGCTAACCAAGTGTGTACCCCTGGGACCTATG 13946	
OY	477 GGTCAAGCTGCCAGACTCGATCCCGACGATATTGAGACACTGGTTAGCCTATTCTGT 536	
Db	13947 CTTCMAATTCCTGAGATGTGAGTTGAGGAGGACATGTGTACCACTTACTGT 14006	
OY	537 TGGCTGCATGCTGTGACGCGCTAGATTCACAGCTGGTGCAGACGCGCTGTTGG 596	
Db	14007 TGGTGTCTGATCTTGTGGCGGAGCTGAGTGTGCTGCTGAAACAAACCTTTTGGTAATGGC 14066	
OY	597 AGGAGTCTATGCGCTTCCACCACTTCTGCTGCAAGGCGCATCATCGGGCAAAAT 656	
Db	14067 AGCTGACACTATGCTGCTTACAAATGTGGCTGCTGGGCTTTCAGTGTGCTAGAT 14126	
OY	657 TGTGTGTCCGACCGCGCTTGCATCAGACAGACTTTGCAAGAACTGGCGCTGAAT 716	
Db	14127 TGTATTGTGAGTGTGATGAGAACCGTTTACCGCTGAGCAAAACAGCTCGCGCATGA 14186	
OY	717 GTTCG 721	
Db	14187 GATTG 14191	
RESULT 15	AE004831 15088 bp DNA linear BCF 30-AUG-2000	
LOCUS	AE004831	
DEFINITION	Pseudomonas aeruginosa PA01, section 392 of 529 of the complete genome	
ACCESSION	AE004831 AE004091	
VERSION	AE004831.1 GI:9950347	
KEYWORDS		
SOURCE	Pseudomonas aeruginosa.	
ORGANISM	Pseudomonas aeruginosa. Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	
REFERENCE	1 (bases 1 to 15088) Stover,C.K., Pham,X.O., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,	
AUTHORS		
TITLE	Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T. Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen	
JOURNAL	Nature 406 (6799), 959-964 (2000)	
MEDLINE	20437337	
PUBMED	10984043	
REFERENCE	2 (bases 1 to 15088)	
AUTHORS	Stover,C.K., Pham,X.O.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reiter,J., Sailer,M.H., Hancock,R.E.W., Lofy,S. and Olson,M.V.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA	
FEATURES	Location/Qualifiers	
source	1..15088	
gene	/organism="Pseudomonas aeruginosa"	
CDS	/strain="PA01"	
	/db_xref="taxon:287"	
	79..1335	
	/gene="PA4142"	
	79..1335	
	/gene="PA4142"	
	/codon_start=1	
	/transl_table=11	
	/product="probable secretion protein"	
	/protein_id="AAG07529.1"	
	/db_xref="GI:9950348"	
	/translation="MRQRLDQNHGCGCEIVYLIPVSTFTLTAAAMALLVGF LFGSTKSTVSQVLPASQVAVHAPQAGIVLRKTVQEQAVRERLNVSSERG SDQPVQAGISRLRLEQRDLRLDQDERDSLSTKVAQLRELTTAAQTD SQQRLALASDAARQVQGLMDQIQQQAELGQROTLOGLERSTLRQOL TERRNELAGISQAQANQLAETRQLSAVEDLSEAKRTLLVTAPESGATVAVLA GQVDSRPLSTVPADTLPQALVAPSKSIGIRPGDAVLIIFYQVAPYQKPCQVYK VQISIRASVSYAELSSMVGCPGLGDDGELVLRVTLDDQAVTAIGQRPRLDSGLL DADIIDTRRLVWLEPLXSLTKL"	
gene	1345..3504	
CDS	/gene="PA4143"	
	1345..3504	
	/gene="PA4143"	
	/codon_start=1	
	/transl_table=11	
	/product="probable toxin transporter"	
	/protein_id="AAG07530.1"	
	/db_xref="GI:9950349"	
	/translation="MAFLDALRLRGLRPLVLTQTEATBGLACIAMIAGYHGHHGL MELRRFVSISGISLKLQITAHRLGCTRAVKLDGLGKIKLPCVIAHNNHNV LKAVDGRGAVLNDPAHGORRLGLEYSRSFTGYALTELPESGFEKQDAPRIKILGLM GKATGLIRSLDAQVLLAGALEVFSLSISPEFLQVTDNVIVSEDRDLSTLAIQGLL LMOQAVSGVAVMMHSTLIGVQWQANVSHLRPAQYFEKRHGDVVRGAVLNA TEEQIVHARQOSHLETVRGIRPLKLFQRODRSRVSGVGLVEQINAGIRQKQLF YQOLNGLFQVENVLVTMGATVMQDQSGVIGTILMAFNAYKQSFDRVSLIDKPEL RMQGLGERLADIVLOAPVSHDILPENLREBEAIEIQGLRYRAEDPVPVLDGL LHAGESAIVTPSCGCKSTLPVLLGLPVEGQIRNAGLDLQGLDGLRELVGT VLQDDVLFAGSLSDNISFPDQPMFWLLQCAOMAIHDDIQAMPGVNTVLVQDMGT LSGGQKQVWLARALYKPRILFLDEATSHLDVHCQDQRYNAIIRALIRIRIVAHPE TIASADRVIVLQCGKVSLESTARLAEQAAAAAEQA"	
gene	3504..4919	
CDS	/gene="PA4144"	
	3504..4919	
	/gene="PA4144"	
	/codon_start=1	
	/transl_table=11	
	/product="probable secretion protein"	
	/protein_id="AAG07531.1"	
	/db_xref="GI:9950350"	
	/translation="KRALGLLGLGLVPGAAAYEPDVEFTTGOVAGQAVYDLGSG	

LPKRGPPPELSELEATEIRLICHDPOTRLAMANKAAQAQVIGKSAVYDRLRLD
ASRGSDMDYDAPVLSGDCRRHRGASLQSWTFEDGSRSAALRNKOILLANAS
ODATLQNTFALAAQVITDALAQAQSLASROYAELIAQNLBAADKTYRAGAALSDRL
QAQTLQAQVDEGALSALNGVIALRMGLADPTLRUGELEAPDVGFAID
EMLAERREHPALAAQALKAASAVESRAQAPSLALSLSLRSHSDAMANGSD
TREARSIGLQNLPLFEGFERTYOVNALARREASEALDTEQOVSLEWNNOST
SVETSLARTRELVSROSLEVOVGRRSSGSMIELLNLMTAVASADQIIRALGN
WQSLRLAASIGRIGFWSLR
5013. .5905
/gene="PA4145"
5015. .5905
/gene="PA4145"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG07532.1"
/db_xref="GI:9950351"
/translation="MIDDIRLVIPAKIAETGVSAGAAALGLSAAATSLHLRLRN
IGRALLYNSRSLTODGTOLDPDRASMLEYKGFEPGRAYSTRNSLRISMPAV
FYNGFTRHIAFIEHPDLDGICADSRHLDIGDIDIRIGDLPDSSLKARHL
LIPROVAPAPFALAHAPLEHPRDLORLEWIGLGRPPSLFRHARDEEVAYDTPR
VRVDSVASCLRLGVGLAAPTYLSDEPIRIGELRAVLPEWRLEPLKYAVAMPNV
PASSIAYTLINLEYAFNPG"
6093. .6497
/gene="PA4146"
6093. .6497
/gene="PA4146"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07533.1"
/db_xref="GI:9950352"
/translation="MARAGEGHAAGSGADPELGIEPLEHLEPGRVPYIQCAWI
RGGDVAHNSDERAGHPQVVEARVGLDRADPEADESQDDAEAGGADERQGP
TMVGASDDGVQVHRTGNNQAQGERAHGEQGS"
complement(6708. .8585)
/gene="acor"
/note="PA4147"
complement(6708. .8585)
/gene="acor"
/codon_start=1
/transl_table=11
/product="transcriptional regulator Acor"
/protein_id="AAG07534.1"
/db_xref="GI:9950353"
/translation="MISASKDHVCVSVLENARLROAPYPTPIIDSMRSMQHR
LDPGSLQGRILSOTLDBECRERSELFLKIASSEVARLGRVADYCYLLADAGRT
IDYRVESTIRNDCKRAGLYLGTQWSEGEQGVAVADRIVYHKNHRAAFID
LTCSPAPEVDPOGELGLVAVSAVSPDQSRQHLIROMVAQSAEIEAFPMHSAOG
HVALRAHRAFGVDSOPVLFAMADGRLALNPAARHLRQLQOVSHIEFVDD
ALHAADAPCLRGAGELHYRLSAPRAKARLREEVDDADLPRIASELSRLAVR
KDRNPVYLQGTGAKVEYAROLHQAASRRKPPVALNCAIPESLISELFGVVG
ATGAAKGRGLDQADGGTLFLDEIGDMPGLGLOTRLRLYAEGBAPLGARQAY
DIOVCATHRDLAALVAAGFREDLYFRIGARFELPLRRESDRILALIRLDEETA
HCGVRIELGEALIECLIGRMPGNVROLHVLRVACALCGATLQADLPAELREBR
TPASACESGGGERDALLDALVRHMKPMAARELIGISRATLYRRVRHGIIMPGQAD
D"
8917. .9717
/gene="PA4148"
8917. .9717
/gene="PA4148"
/codon_start=1
/transl_table=11
/product="probable short-chain dehydrogenase"
/protein_id="AAG07535.1"
/db_xref="GI:9950354"
/translation="MHATSSDFSLSSRYALVTGAGRGIGRGITALLARAGADVAD
LPOVAAETFAAARSIGRSIALGVDSGDVSRAMAVEVATEFGSLDVAVNNAGTIS
IKVAELSLADMDRWNNVAVRGFLCCQELPLMOQRMGRILVNLSSIGKGLPLIA
HYCAKFAVIGFSNLAKEVARGVTVNALGIVGTGWMRGEDDLSGRWRQAGESEA
QSMERHQASLPOGEQVYEDMGQLVTVIACAPRHTYGAALAVDGGESL"
9714. .10734
/gene="PA4149"
9714. .10734
/gene="PA4149"

9714. .10734
/gene="PA4149"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07536.1"
/db_xref="GI:9950355"
/translation="MTPSPVIGIIANPASGRDLRLTAHAGLSHSDKASVORLLA
ARGATGTHALLPDMGIAAVALRASOTROARGHMPLEFLMPLRSOVEDRLA
RMRVERGVALTAVIGDSTHKAVAAEVDVLLSTGCTNNAPELRAGAGLGL
CASNVPEPGLRKRRLVREPRGICEMALVAVASPOPTARASREDLAEYF
VSFAEPHAGLSALGLMSPVRSREPHGAMRLHPDAESLLVPLAPLGVCSAA
GYLQGVHAHAPLSLSTGLALDGERIEERSAIDRPSITLDPSPESVDVPAFLAAYAR
RLAGQRTPEMP"
10777. .11751
/gene="PA4150"
10777. .11751
/gene="PA4150"
/codon_start=1
/transl_table=11
/product="probable dehydrogenase E1 component"
/protein_id="AAG07537.1"
/db_xref="GI:9950356"
/translation="MSTLSTQDLHAYRVMTIRAFERLHVEATGEIPGFVHLVAG
EASAGVMAHLRDDCIASIRHGHIAQGVYHGMMAEYKKTGVCQKGSMSH
TADLEKMLGANGLYAGAPLAALAKLKGDAVAVAFEGDGSNEGAVFAMNL
ASVNLPLGVANNVGAATAAANSVACDHIADRAAGFEMGTVYDGFPAVHEAA
GATIRARAGBPSLIEVKLTRYIGHFEGDAQYRDPEVYHRETRDCLKQFRERTC
HAGLSASDLDAIDAEVARIEDAVORAKNDKPEPDLDDVYVSYF"
11784. .12803
/gene="acob"
/note="PA4151"
11784. .12803
/gene="acob"
/codon_start=1
/transl_table=11
/product="acetoin catabolism protein Acob"
/protein_id="AAG07538.1"
/db_xref="GI:9950357"
/translation="MARKISYQALINELAQEMRDPVFLIMGEDMAGAGAPGEDDA

Query Match 5.6%; Score 63.8; DB 1; Length 15086;
Best Local Similarity 46.5%; Pred. No. 1.5e-06;
Matches 326; Conservative 0; Mismatches 357; Indels 18; Gaps 3;

QY 432 CAGCGCGGTTTCCCGATAGCTGCTTACGGGTGAGCACACATGTCAGCTGCACA 491
DB 14330 CAACGGCGCTTCGCCGAATAGCTACAGTACCGGCCAACCTGCTTACGCCCTGCCGC 14389
QY 492 CTCGATTCGCCGACATATTGAGCACTGGTGAACCTATTCTGTTGCTGGCATGCTGT 551
DB 14390 CGGATTTCCGAGGAGCGCGCGCGCTGATCGAGCGGCTGGCGGTGGCATGCACGCGGT 14449
QY 552 TGAACCGCTAGATTCCAGCTGCTGAGACGGCCCTGCTTCTTGAGAGAGTCTCTATCGG 611
DB 14450 GAAAGAGCGCGGAGCGCTCTCGGGGAGAACGTTGGTGGCGCGGACCACTCGG 14509
QY 612 CCTTGCACCATCTTGTCTGTCTGCAAGCATCATGCGGCAAAATGTGTCTTCGAGCC 671
DB 14510 CCTGACACCATCATGTGCGCGCGCGCGCGCGCGCGGACACAGTGTATGCCCTGAGAT 14569
QY 672 GGCCTTGATCAGAAGACAGTTTGCAAAAGAACTGGCGCTGAAGTGTTCGATCTTCTAC 731
DB 14570 GTCTTCGGCGGCGAAGCGCAAGGCGCTGAGAGTGGCGCGCACAGCAAGTCTGACCCCTC 14629
QY 732 ATGTGACGACGCAAAATGCTGTCTCAAGCTAATGCTGCGGAGAACGAGGATTCATGC 791
DB 14630 GCGGTCGAGCGCCCTAGCGGAATTCGCGC---CTTACCGCGCGGCTTGGCGCGCACGT 14686
QY 792 AGCTTCGACGTGCTGTGTTCTCTCAGACATTCACCACTCAATGTGCGCACGCGGAC 851
DB 14687 CAGCTTCGATGATGATGGGCAACAAGACACACCGCAAGCTGCGCATGATGATCGCA 14746


```

OY 852 TTCTGGAAATCGCCGTCAATGTGCGCCCTTTGGGGAGACCACCACAATTGGATTTCATGCCAAT 911
Db 14747 GCGCCGCAAAATGCGTGTGTTGGGAATCT---TCGAGGAGGCCCACTGAATTCATTTCTT 14803
OY 912 GTCTGTGACTTACGAGGAAATACGCTACCGGCTCCATGTGCTACACCGTCAAGGACTT 971
Db 14804 CGAGCTGGTTCCACCGAAGACAGTTGCTCGGGGCCCTCGCTACACGCGGAGTTGCG 14863
OY 972 CCAGGAAGTTGTCAAGGCTTGAAGATGTCATATCTTTGGACAAGCGCGCAAGAT 1031
Db 14864 CGAGCTGATCCGCTTCATAGCCGAGCGCCGCT-----CGACATCGCACCGCT 14911
OY 1032 GATTACAGGCAAAAGTCCACTAAAGGAGCGAGTCGAGAAGGCTTTAAACAGCTGATCGA 1091
Db 14912 GGTGACCGGTGCAATCGGCTGGAAGAAATGTCGAGCGCGTTTCGAAGAGTTGTTGAA 14971
OY 1092 GCACAAGGAGAACAATGTCAAGATCCTGCTGACGCCGCAACG 1132
Db 14972 CACACAAGGACACACGATGAAAGATCATGCTCGCCGGGCG 15012

```

Search completed: May 2, 2003, 09:10:20
 Job time : 3438 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 06:47:36 ; Search time 299 Seconds

(without alignments)
8608.810 Million cell updates/sec

Title: US-10-020-674-1

Perfect score: 1143
Sequence: 1 atgaaggttacttactta.....cgccgaacgaggttctctaa 1143

Scoring table: IDENTITY_NUC

Gapop 10.0, Capext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.6	8.2	536	21	AAF08292
2	83	7.3	1041	24	AA028420
3	69.8	6.1	339	24	ABR77781
4	67	5.9	1113	21	AA049213
5	67	5.9	1316	21	AA049216
6	67	5.9	1347	21	AA036132
7	67	5.9	1347	21	AA049218
8	67	5.9	1413	21	AA049219
9	65.4	5.7	1315	21	AA038499

10	65.4	5.7	1407	21	AA033156	Arabidopsis thalia
11	63.8	5.6	1092	23	AA054244	Pseudomonas aerugi
12	62.8	5.5	1344	23	AB003803	Drosophila melanog
13	62.2	5.4	1338	23	AB007985	Drosophila melanog
14	57.4	5.0	933	22	AA043317	T. aestivum orbilo
15	56.6	5.0	1163020	24	AB067197	Listeria innocua C
16	56.6	5.0	3011208	24	AA069245	Listeria innocua D
17	56.4	4.9	589	21	AA067017	Pinus radiata tran
18	56.4	4.9	589	21	AA067331	Pinus radiata D-xy
19	56.4	4.9	590	21	AA067340	Pinus radiata D-xy
20	55.4	4.8	1038	24	AA033483	Human drug metabol
21	52.8	4.6	4164	23	AB003802	Drosophila melanog
22	52.2	4.6	2410	21	AA015951	Human prostate can
23	52.2	4.6	2519	24	AB095651	Gene #2189 used to
24	52.2	4.6	2519	24	AB062665	Colon adenocarcino
25	52.2	4.6	2519	24	AB069415	Prostate cancer re
26	52.2	4.6	2597	23	AB073357	Human prostate exp
27	52.2	4.6	2597	23	AB073354	Human prostate exp
28	52.2	4.6	2597	23	AB073354	Human prostate exp
29	51.8	4.5	1301	24	AB070384	Listeria monocytog
30	51.4	4.5	534	21	AA067346	Pinus radiata D-xy
31	50.2	4.4	2944528	24	AA03041	Listeria monocytog
32	50	4.4	484	21	AA067329	Eucalyptus grandis
33	49.4	4.3	872	22	AA052673	S. epidermidis ope
34	49.4	4.3	1065	24	AB091074	Staphylococcus epi
35	49.4	4.3	4019	22	AA054846	S. epidermidis gen
36	49	4.3	1610	22	AA043316	G. max sorbitol de
37	49	4.3	2703	23	AA052993	DNA encoding novel
38	47.2	4.1	2774	15	AA066579	Sorbitol dehydroge
39	46.6	4.1	1302	24	AB068679	Listeria monocytog
40	46	4.0	461	21	AA056883	Pinus radiata tran
41	45.8	4.0	1014	24	AB073260	Bacillus lichenifo
42	45.8	4.0	3866	23	AB070984	Drosophila melanog
43	45.4	4.0	1053	21	AA082824	Fusarium venenatum
44	45	3.9	513	21	AA077873	CDNA encoding huma
45	45	3.9	513	22	AA128611	Colon tumour relat

ALIGNMENTS

RESULT 1
AAF08292 standard; CDNA; 536 BP.
XX
AC AAF08292:
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:815.
XX
XX
Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
DR WPI; 2000-594572/56.

RESULT 3
ABN77781
ID ABN77781 standard; cDNA; 339 BP
XY

[illegible]

PN WO200190366-A2

24-MAY-2001; 2001WO-US17076.

PR 24-MAY-2000; 2000US-206690P.

(CURA-) CURAGEN CORP.

Leach MD, Shlmkets RA;

DR WP1; 2002-106200/14
DR B-PENB; APN337EE

Novel human poliovirus

preventing and treating cardiovascular disease, neurodegenerative, hyperlipidemic disorders and Alzheimer's disease.

transpiration -

Created: 1; Page 1013; 2508pp; English

sequence ABP1028-ABP35561 represent 4334 novel human proteins designated ORF (open reading frame) 1-4334, and sequences AN75054-ABN7937 represent at least 806 identical them. The invention also encompasses polypeptides at least 80% identical to the ORF1-ORF4334 (collectively the ORF1 to 43 ORF43 proteins), polynucleotides at least 85% identical to the ORF1 to 43 ORF43 sequences, vectors and host cells comprising ORF1 to 43 polynucleotides, the recombinant production of ORF1 proteins, antibodies specific for ORF1 proteins, methods of detecting ORF1 polynucleotides and

CC polypeptides methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, homeostatic activity, thrombolytic activity,
CC receptor/ligand, anti-inflammatory activity, tumour inhibition activity,
CC and antitumorigenic activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX

Sequence 339 BP; 78 A; 70 C; 101 G; 90 T; 0 other;

Sequence 339 BP; 78 A; 70 C; 101 G; 90 T; 0 other,

Query Match	6.18;	Score 69.8;	DB 24;	Length 339
-------------	-------	-------------	--------	------------

Matches 170; Conservative . 0; Mismatches 167; Indels 0; Gaps 0;

254 GCTCTGGTGCACAGTGTGAACCTGGTGACAGAGTCCGAGTTGAAGCTACGTCGATT 31

Db 2 GCCCTGGTGAACCTTGAAGTGGGAGACAAGGTAGTTGTCGAGCCACAGGTACAT 61

314 GCTCCGACAGATCGCGCTACAAGGACACGGTCCGCCAAGACCTTGGGCTCTGTATGGCCT 37

DB 62 GCAGAGACCGGTATCGTTGCCCCCTGTGCCCAACGTTGACCAAGGATGTCGCTGCTT 12

3/4 GCCAGAGCGGATCTCCGAAGTCTGTGCGCTGAGCTTCTGCGGTTTGGGTGTCCTCA 43

LD 122 GCATAAAGGGCTACCTATACACTTTGTTCATATTTGGGGCTTTGTGGTGCGGGTGTGCACA 18

434 GCGGCGTCTTGGCCGAGTACGTCGTTACGGTGAGGACCACATGCTCAAGCTGCCAGACT 49

102 GCGGAGATTTGCAAGACGTTGTGATGAACGAATCTCACTGCTACAAAGTACCGACT 24

434 CGAII CCGACGATATTTGGAGCACITGGITTGAGCCCTATTTCTGTGCCCTGGCATGCTGTTG 55

22 242 1CGGCCCCAGAGCTTTGATTTCAACCGTTGGCTGTGTGCTGGCATGCAATTA 30

204 AACCCCAAGATTCAGCCCTGGT 590

[illegible]

RESULT 4

ID AAC49213 standard; DNA; 1113 BP.

AC AAC49213;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 60340.

KW Hybridisation assay; genetic mapping; gene expression control;

metabolic pathway; promoter; termination sequence; ss..

05 *Arabidopsis thaliana*.

XX EP103405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0125264.
 PR 29-MAR-1999; 99US-0126783.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 28-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 04-MAY-1999; 99US-0132407.
 PR 05-MAY-1999; 99US-0132484.
 PR 06-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 18-MAY-1999; 99US-0134370.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135533.
 PR 24-MAY-1999; 99US-0135529.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140635.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140981.
 PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144886.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148315.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.

```

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155179.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161930.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 5.9%; Score 67; DB 21; Length 1113;
 Best Local Similarity 47.1%; Pred. No. 8.2e-10;
 Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

```

OY 357 TGGCCTGTATGGCCTGCCAGACGGATCTCCGACTGCTGCTGAGCTCTG 416
DB 345 TTGGAGATGCAATCTCTGACGAGGAGCATACACCTTTGTCAGAAATGAGTTCTT 404
OY 417 CGGTTTGGGTGTCACAGCGGGGTTTTCGCACTACGTCGTTTACGGTGAAGCACAT 476
DB 405 TGCACCTCCACCGGTTTATGCTTTAGCTAACCAAGTGTATACCTCGGACATATG 464
OY 477 GGTCAAGCTGCACACTGATTCCTCCAGCATATGAGCACTGGTGGCTATTTCTGT 536
DB 465 CTTCAAATGCGCTGAGATGATGATTTGGAGGAGCAATGTTGTAACCACTTACTGT 524
OY 537 TGCTTGCATGCTTTACAGCGCTGATTCACAGCTGTGTCAGACGGCCCTGTTCTTGG 596
DB 525 TGGTGTCCATGCTTGTGCGGAGCTGAGTGTGCTGAAACAAACCTTTGGTAATGGG 584
OY 597 AGAGGTCCTATGCGCTTGCACCACTTCTTGCTCTCCAGAGCCATATCGGGCAAAAT 656
DB 585 AGCGGACCTATTTGCTTTTACAAATGTTGGCTGCTCGGCTTTAGTGTGCTTGAAT 644
OY 657 TGTGTGTTCCGACCGGCTTGTATCAGAACAGATTTCAGAAAGAACTGGCGCTGAGT 716
DB 645 TGTATTTGTGATGTGATGAGAACCGTTTAGCCGTAGCAAAACAGCTGCGGCAATGA 704
OY 717 GTTCGATTCCTTACATGTGACAGCAAGCAATGCTGTTTCTCAGCGTATGTTGCGGAGAA 776
DB 705 GATTGTACAAGTACAAACAACTTAAAGAGTGTGTTGAGAGTTGTAACAGATTGAGAA 764

```

```

OY 777 CG-----AGGATTCATGACGCTTGCAGTCTGTGTTCTCTCAGACATTCAC 827
DB 765 AGCTATGGGTCACAAACATGATGTAACATTCGACGTGCGGTTTACAAACATGTC 824
OY 828 CACCTCAATTTGCGCCAGGAGGACCTTCTGATGATGCCCTCATGTGCG 874
DB 825 GACACATTCAGACGCACTGTTGGGGGTAAAGTGTGCTGTGCG 871

RESULT 5
AAC49216
ID AAC49216 standard; DNA, 1316 BP.
XX
XX AAC49216;
XX
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 60351.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.

```


Db 433 TTGAGATGCAMTCTCTGACGGAGGACGATACACACTTTGTCCAGAAATGATTTCT 492
Oy 417 CGTTTGGGTGGTGGCCAGGGCGTTTCCAGTACGTCGTTTACGTGAGACCAT 476
Db 493 TGCACCTCCACCGGTTTCATGCTTCTTACCTAACCAAGTGATACACCTCGGACCTATG 552
Oy 477 GGTCAAGCTCCAGACCTCCATCCGACGATATTGGAGACGAGTGGACCTATTTCTGT 536
Db 553 CTTCAAAATGCGCGAATATGTAGTTTGAGAGAGACGAATGTGAAACCACTTAGTGT 612
Oy 537 TGCCTGGCATGCTGTTGAAAGCGGCTAGATTCAGCCTGTGACAGCGGCTGTCTTGG 596
Db 613 TGTGTCCATGCTTGTGCGCGGACGAGCTGTGTCTTAAACAAACGTTTGTATGGG 672
Oy 597 AGGAGTCTATCGGCTTGCACCATCTTGTCTGCAAGGCCATCATGCGGCAAAAT 656
Db 673 ACCTGACCTATTTGGCTTGTTCATATGTTGCTGCTCGGCTTTCAGTGTCTTCAAGAT 732
Oy 657 TGTGTGTTCCGAGCGGCTTGTATCAGAACAGACATTTGCAAGAACTGGCGCTGAAGT 716
Db 733 TGTATTTGATGATGTTGATGAGAACCGTTTACCGTAGCAAAACGCTGCGGACAGATGA 792
Oy 717 GTTCATCTTCTACATGAGACGACGACCAATGCTTCTCAAGGCTATGCTCCGAGAA 776
Db 793 GATTGTACAGTACACAACTTAGAGAGATGTTGTTCTGAGGTTGAACAGATTTCAGAA 852
Oy 777 CG 778
Db 853 AG 854

RESULT 6
AAC36132
ID AAC36132 standard; DNA; 1347 BP.
AC36132;
XX
AC 17-OCT-2000 (first entry)
XX
DT Arabidopsis thaliana DNA fragment seq ID NO: 12661.
XX
DE
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123380.
PR 09-MAR-1999; 99US-0123358.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128224.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139432.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147312.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139
PR 23-SEP-1999; 99US-0155486
PR 24-SEP-1999; 99US-0155659
PR 28-SEP-1999; 99US-0156458
PR 29-SEP-1999; 99US-0156596
PR 04-OCT-1999; 99US-0157117
PR 05-OCT-1999; 99US-0157753
PR 06-OCT-1999; 99US-0157865
PR 07-OCT-1999; 99US-0158029
PR 08-OCT-1999; 99US-0158232
PR 12-OCT-1999; 99US-0158369
PR 13-OCT-1999; 99US-0159293
PR 13-OCT-1999; 99US-0159295
PR 14-OCT-1999; 99US-0159329
PR 14-OCT-1999; 99US-0159330
PR 14-OCT-1999; 99US-0159331
PR 14-OCT-1999; 99US-0159637
PR 14-OCT-1999; 99US-0159638
PR 18-OCT-1999; 99US-0159584
PR 21-OCT-1999; 99US-0160741
PR 21-OCT-1999; 99US-0160767
PR 21-OCT-1999; 99US-0160768
PR 21-OCT-1999; 99US-0160770
PR 21-OCT-1999; 99US-0160814
PR 21-OCT-1999; 99US-0160815
PR 22-OCT-1999; 99US-0160980
PR 22-OCT-1999; 99US-0160981
PR 22-OCT-1999; 99US-0160989
PR 25-OCT-1999; 99US-0161404
PR 25-OCT-1999; 99US-0161405
PR 25-OCT-1999; 99US-0161406
PR 26-OCT-1999; 99US-0161359
PR 26-OCT-1999; 99US-0161360
PR 26-OCT-1999; 99US-0161361
PR 28-OCT-1999; 99US-0161920
PR 28-OCT-1999; 99US-0161922
PR 28-OCT-1999; 99US-0161992
PR 29-OCT-1999; 99US-0162142

Query Match 5.98; Score 67; DB 21; Length 1347;
Best Local Similarity 47.18; Pred. No. 9.1e-10;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;

QY 357 TGGGCTCTGATGCGCTGCAGACGAGATCTCGAACTGCTGCGCTGAGCTTCTG 416
DB 464 TTGGAGATGCAATCTCTGCGAGGAGACATACAACTTGTGCGAAATGAGTTCTT 523
QY 417 CGGTTGGGTGTGCCAGCGCGGGTTTGGCGAGTACGTCTTACGGTGAAGACACAT 476
DB 524 TGCACCTCCACCGGCTTCATCGCTCTTGTAGCTAACCAAGTGTACACCTCGGACCTATG 583
QY 477 GGTCAAGCTCCAGACTCGATTCGCCAGCATATTGGAGCGACTGTTGAGCTATTCTGT 536
DB 584 CTTCAAATTCCTGAGAAATGTGATTGGAGAAAGAGCAATGTGTAACCACTTATGTGT 643
QY 537 TGCCCTGGCATCTGTTGAAGCCCTAGATTCAGCCCTGTCAGACGCGCCCTGTTCTTG 596
DB 644 TGGTGTCCATGCTTGTGCGCGACTAGTGTGCTCCGTAACAAAGCTTTGGTAATGG 703
QY 597 AGAGAGTCTATGCGCCCTGCCACCATTTCTGCTCTGCAAGGCCATCATCGGAGAAAT 656
DB 704 AGCTGACCTATTGGCTTGTACAAATGTTGCTGCTCGGGCTTTCAGTGTGCTAGAAAT 763
QY 657 TGTGTGTCCGAGCCGCTTGTATCAGAAAGAGCTTTGCAAGAGAACTGGGCGCTGAGT 716
DB 764 TGTATTGTGGATGTGATGAGAACCGTTTACCGTAGCAAAACAGTCGCGCAGATGA 823
QY 717 GTTCGATCTTTCATGTCAGCAGCAAGCAATGCTGTTCTCAAGGCTATGCTGCGGAGAA 776
DB 824 GATGTCAAGTGCACAAACACTTAGAGATGTGTTGCTGAGGTTCAACAGATTCGAA 883
QY 777 CG-----AGGATTCATGACGACCTTCGACTGCTGTGTTGCTCAGACATTCAC 827

DB 884 AGCTATGGGGTCAACATGATGTATACATTCGACTGTGCGGGTTTAAACAAACCATGTC 943
QY 828 CACCTCATTTGCGCCACGCGGACCTTTGGAATCCCGCATATGCG 874
DB 944 GACAGCATTTAGCAGCCACTGCTTGTGCGGTAAGTCTGTCTTGTGCG 990

RESULT 8
AAC49219
ID AAC49219 standard; DNA; 1413 BP.

AAC49219;

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 60363.

Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135829.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

28-MAY-1999; 99US-0136782.

01-JUN-1999; 99US-0137222.

03-JUN-1999; 99US-0137528.

04-JUN-1999; 99US-0137502.

07-JUN-1999; 99US-0137724.

08-JUN-1999; 99US-0138094.

10-JUN-1999; 99US-0138540.

14-JUN-1999; 99US-0138847.

14-JUN-1999; 99US-0139119.

```
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149176.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155132.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.9%; Score 67; DB 21; Length 1413;
Best Local Similarity 47.1%; Pred. No. 9.3e-10;
Matches 248; Conservative 0; Mismatches 270; Indels 9; Gaps 1;
```


PR	21-OCT-1999;	990S-0160770.
PR	21-OCT-1999;	990S-0160814.
PR	21-OCT-1999;	990S-0160815.
PR	22-OCT-1999;	990S-0160980.
PR	22-OCT-1999;	990S-0160981.
PR	22-OCT-1999;	990S-0160989.
PR	25-OCT-1999;	990S-0161404.
PR	25-OCT-1999;	990S-0161405.
PR	25-OCT-1999;	990S-0161406.
PR	26-OCT-1999;	990S-0161359.
PR	26-OCT-1999;	990S-0161360.
PR	26-OCT-1999;	990S-0161361.
PR	28-OCT-1999;	990S-0161920.
PR	28-OCT-1999;	990S-0161992.
PR	28-OCT-1999;	990S-0161993.
PR	29-OCT-1999;	990S-0162142.

Query Match	5.7%	Score 65.4;	DB 21;	Length 1315;
Best Local Similarity	47.2%	Pred. No. 2.8e-09;		
Matches 199; Conservative	0;	Mismatches 223;	Indels 0;	Gaps 0

```

XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0122180.
PR 09-MAR-1999; 99US-0122544.
PR 23-MAR-1999; 99US-0122788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0128262.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134368.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135529.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139432.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147938.
PR 10-AUG-1999; 99US-0148177.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

```


PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161982.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 5.7%; Score 65.4; DB 21; Length 1407;

Best Local Similarity 46.9%; Pred. No. 2,9e-09;
Matches 247; Conservative 0; Mismatches 271; Indels 9; Gaps 1;

QY 357 TGGCTCTGTATGGCTGCCAGACGATTCGCAATGCTGTGCTGCTAGCTTCTG 416
DB 522 TTGGAGATGCAATCTCTGCAGAGAAAGACGATACCACTATGTCAGAAATGTTCTT 581
QY 417 CGGTTGGGTGGTCCAGCGCGGCTTTGCCAGTACGCTGTTACGGTGAACACAT 476
DB 582 TGCACCTCCACCGGTTTACGCTCTTTAGCTAACGAGTGGTTCCTCGCGATCTATG 641
QY 477 GGTCAAGCTGCCAGACGATTCGCCAGATATTGAGCAGTGGTTCCTATTTCTGT 536
DB 642 CTTCAAATGGCTGAGATGATGATTTGAGAGAAAGACATGTGTGAACACATTAGTGT 701
QY 537 TGCCTGGCAGCTGTGAACGCGCTAGATTCAGCCTGGTGAAGCGGCTGTCTTGG 596
DB 702 TGGTGTCAATCTGTGCGCGAGCTTAGCTGTGTCCTGAACAACAGTTTGGTAATGG 761
QY 597 AGGAGTCTCTATCGGCTTGCACCATTTCTGTGTAAGCCATCATTCGGGCAATAT 656
DB 762 AGCTGAGCTATTTGGCTTTGATACATGTTGGCTGCGGCTTTCAGTGTGCTAGAT 821
QY 657 TGTGTGTTCCAGCGCGCTTGTATCAGAAAGACAGTTTGCAGAAAGAGTGGCGTGAAGT 716
DB 822 TGTATTTGTGATGTTGATGAGAACCGTTTACCGCTAGCAAAACGCTGGCGCAGATGA 881
QY 717 GTTCGATCTCTTACATGATGAGCAGCAATGCTTTCTCAAGCTATGTCGCGAGAA 776
DB 882 GATGTGACAGTGAACAACACTTAGAGAGATGTTGGTCTGAGGTGAACGATTCAGAA 941
QY 777 CG-----AGGATTTCCATGCAAGCTTGCAGCTGCTGTGTTCTCTGACATTTAC 827
DB 942 AGCTATGGGTGCAAAACATGATGATACATTCGACTGTGCGGGTTTAAACAAACCATGTC 1001

OY 828 CACCTCATTTGCGCCAGCGACCTTTCGATGCGCCGTCATGNG 874
DB 1002 GACACATTTAGCAGCCACTCGTTGTGGCGGTAAGTGTGTCTGCG 1048

RESULT 11

AAS54244
ID AAS54244 standard; DNA; 1092 BP.

AC AAS54244;

DT 13-FEB-2002 (first entry)

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #375.

KW Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

XX WO200170955-A2.

PD 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-26308P.

XX (ELIT-) ELITRA PHARM INC.

XX PA

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI; 2001-611495/70.

XX P-PSDB; AAU36385.

PT New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

PS Claim 27; Seq ID No 7881; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in

CC a wide variety of organisms. The present sequence encodes an

CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.

CC

SO Sequence 1092 BP; 191 A; 380 C; 352 G; 169 T; 0 other;

Query Match 5.6%; Score 63.8; DB 23; Length 1092;
Best Local Similarity 46.5%; Pred. No. 7.8e-09;
Matches 326; Conservative 0; Mismatches 357; Indels 18; Gaps 3;

```

OY 432 CAGCGCGGTTTGGCGATGCTGTTACGCTGAGACACACATGTCAGTGCACAA 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 405 CACGCGCGGCTTCCGGAATGCTCAACGTACCGCCACCTGCTACGCCCTGCG 464
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 492 CTCGATTCCTGAGATATTGAGACATGGTGGTGAAGCTATTTTGTGCTGCGATCTGT 551
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 465 CGGATTTCCAGAGGAGCGCGCGCGCTGATCGAGCCGCTGGCGGTGGGATGACCGGGT 524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 552 TGAAGCGCTAGATTCAGCTGCTGAGACAGCGCCCTGTTTGGAGAGTCTTATCGG 611
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 525 GAAGAGCGCGGAGCTCTCGGCAACGAGTGTGTGTGCGCGCCGACCATCGG 584
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 612 CCTGCGACCATCTGCTGCTGCTGCAAGGCATATATGCGGCAAAATGTGTGCTCCAGCC 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 585 CCTGAGACCATCATGATGCGCGCGCGCGCGCGCGGACACAGTATGCGCTGAGAT 644
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 672 GCGCTTGATCAGAGACAGTTCGCAAGAGACTGGCGCTGAGTGTGATCTTCTAC 731
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 645 GTCCTCGGCGGCAAGGCGCAAGGCGCTGAGAGTGGCGGCAAGCATGCTGACCCCTC 704
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 732 ATGTGACGACGCAATGCTGTTCTCAAGCTATGTCGCGAGACAGAGGATTCATGC 791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 705 GCGGTGCGAGCGCTTGGGCAATCCGCGC---CTCACCGCGCGCTTGGCGCCGACGT 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 792 AGCCTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 762 CAGCTTGAATGATGATGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 852 TTCTGGAATCGCGCTCAATGTGGCCCTTGGGAGACACCCCAATGATTCATGCAAT 911
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 822 GCGCGGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 912 GTCCTGACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 971
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 879 CGAGCTGCTTCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 972 CCAGGAGTGTTCAGAGCGCTTGGAGAGATGCTCATATCTTTGAGACAAAGCGCGCAAGAT 1031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 939 CGAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1032 GATTACAGCAAGTCCACCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1091
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 987 GGTGACCGGTGCGATGCGGTGAGAAATGTCGAGCGCGGCTTGGAGAGAGAGAGAGAG 1046
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1092 GCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1047 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 12

ABL03803
ID ABL03803 standard; cDNA; 1344 BP.

AC ABL03803;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5891.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

```

XX (PEKE ) PE CORP NY.
PA Venter JC, Adams M, Li PMD, Myers EW;
PI WPI; 2001-656860/75.
DR P-PSDB; ABB59700.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PS interactions.
XX Claim 1: SEQ ID NO 5891; 21bp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB17737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_sequences.
XX
XX Sequence 1344 BP; 319 A; 364 C; 367 G; 294 T; 0 other;

```

Query Match 5.5%; Score 62.8; DB 23; Length 1344;
Best Local Similarity 49.1%; Pred. No. 1.8e-08;
Matches 166; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

```

OY 373 TGGCAGAGGGGATCTCCGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 432
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 565 TGCACGACGAGGCAATACAACTGTGCCCCGGAATGCTTGTGTCACAGCCGCCCTAC 624
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 433 AGCGCGGCTTTCGCGAGTACGCTGTTACGGTGAAGACACATGCTCAAGCTGCAGAC 492
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 625 GACGCGAACCTCACCGCTACATGAGCAGTGGCGGCACTTCTGCTTCAAGCTGCGCGGAC 684
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 493 TCGATTCGCGAGCATTTGAGAGCATGCTGAGCCATTTGCTGCTGCGCATGCTGTT 552
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 685 CACGTACCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 744
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 553 GAACGCGTATGATTTCCAGCTGCTGTCAGAGCGCCCTGTTTGGAGAGAGTCTTACGCGC 612
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 745 AAGCGGCGGAGGTACACCTGGGCTCAAAAAGTTCTATCTGTGGAGACAGACCATTTGCG 804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 613 CTGGCCACCATTTCTGCTGTCAGAGCGCATATGCGGCGCAAAATTTGTGTTCCAGCGC 672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 805 CTGTGCTACTTTATGCTGCTCTCAAGCCATGAGGCTGTGAGATCCTCATTTACGATCTT 864
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 673 GCCTTGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 710
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 865 GTGCAGCAACGCGCTGATGATGATTAAGAGACTTAAAGTGTGC 902
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 13

ABL07985
ID ABL07985 standard; cDNA; 1338 BP.

AC ABL07985;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18437.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

QY 659 TGTGTTCCGACGGCCTTGATCAGAGACAGTTTCAGAGAACTGGCGCTGAAGTCT 718
 DB 609 TCATCGCCGACGTGACGACACCGCCTCTCCGTGCGCAAGTCCCTCGGCGGAGCGCG 668
 QY 719 TCG 721
 DB 669 TCG 671

RESULT 15
 ID ABO67197 standard; DNA; 1163020 BP.
 AC ABO67197;
 DT 29-AUG-2002 (first entry)
 DE Listeria innocua contig DNA sequence #10.
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KM Infection; ds.
 OS Listeria innocua.
 PN WO200228891-A2.
 PD 11-APR-2002.
 PP 04-OCT-2001; 2001WO-FR03061.
 PR 04-OCT-2000; 2000FR-0012697.
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Kunst F, Glaser P;
 DR WPI; 2002-332479/37.
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 PS Claim 5; SEQ ID 10; 180pp; French.
 XX The present invention relates to nucleic acid sequences
 CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1163020 BP; 388339 A; 197589 C; 235612 G; 341474 T; 6 other;

Query Match 5.0%; Score 56.6; DB 24; Length 1163020;
 Best Local Similarity 48.3%; Pred. No. 5.4e-05;
 Matches 158; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

QY 435 CGGCGGTTTGGCCGACGTGCTTTACGGTGAGAGACACATGCTCAAGCTGCCAGACTC 494
 DB 437437 CGAGAGCTATGCTGATATATCAAACTTCACACAAAATGTAGTAAATATCCAGCGA 437496
 QY 495 GATTCGCGAGATTTGGAGCACTGCTGATTTCTGTTGCTGGCATGCTGTGA 554
 DB 437497 ATTAGATTATGAGCGCGCTGCACTTCTGAGCTGTGATGATGATTTTA 437556
 QY 555 ACGCGCTAGATTCCAGCTGTGTCAGAGCGCCTGTTCTTGGAGAGATCTATCGGCCT 614

DB 437557 TCATACGAATTTACAGCTGGCGATGATGTCTGTGCGAAGCGCAATATCGGTTT 437616
 QY 615 TGCCACCATTTCTTGTCTGCAAGGCCATCATGCGGCAAAATGTGTTCGAGCCGCG 674
 DB 437617 GCTAGCAATTCATGCGCAAAAGTTTGGCGCAAGAAGATTATTCGATGTGATGA 437676
 QY 675 CTTGATCAGAAACAGATTTCGAAAGAACCTGGGCGTGAAGTGTTCGATCCTTCATG 734
 DB 437677 CGATAAAAAGTTAGCTCTTGGCAAGAGTGGAGACAGACGTACTATATTTCTTAA 437736
 QY 735 TGACGACGCAAAATGCTGTTCTCAAGGC 761
 DB 437737 AGAAGATCCATTAGAAAGTAGTCGACG 437763

Search completed: May 2, 2003, 08:24:55
 Job time : 1025 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 08:02:56 ; Search time 1611 Seconds
(without alignments) 11490.653 Million cell updates/sec

Title: US-10-020-674-1

Perfect score: 1143
Sequence: 1 atgaaggcttacttacta.....cgccgaacgaggttctctaa 1143

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

EST:
1: em_estb1:
2: em_estb2:
3: em_estb3:
4: em_estb4:
5: em_estb5:
6: em_estb6:
7: em_estb7:
8: em_estb8:
9: em_estb9:
10: em_estb10:
11: em_estb11:
12: em_estb12:
13: em_estb13:
14: em_estb14:
15: em_estb15:
16: em_estb16:
17: em_estb17:
18: em_estb18:
19: em_estb19:
20: em_estb20:
21: em_estb21:
22: em_estb22:
23: em_estb23:
24: em_estb24:
25: em_estb25:
26: em_estb26:
27: em_estb27:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	735.2	64.3	1057	17	CNS0778D	AL432371 T3 end of
2	240.2	21.0	893	17	CNS07444	AL428330 clone BAO
3	233.4	20.4	946	17	CNS06Y2B	AL421677 T3 end of
4	195	17.1	1066	17	CNS06KBW	AL402890 T7 end of
5	184	16.1	976	17	CNS06L93	AL403885 T7 end of
6	159.2	13.9	994	17	CNS06XOA	AL420056 T3 end of

7	148.8	13.0	923	17	CNS073DD	AL427367 clone BAO
8	120.4	10.5	908	17	CNS07443	AL428330 clone BAO
9	107.4	9.4	924	17	CNS070R5	AL423975 T3 end of
10	100.2	8.8	1045	17	CNS0760Q	AL431664 T7 end of
11	99.8	8.7	982	17	CNS06L5E	AL403896 T7 end of
12	97.6	8.5	1105	17	CNS07E9Y	AL441500 T3 end of
13	94.4	8.3	489	17	A2931833	A2931833 474 dhz90
14	77	6.7	493	9	AL696793	AL696793
15	62.6	5.5	596	12	BF717166	BF717166 Lf74 near
16	62	5.4	673	13	BM513099	BM513099 Kx72F07.Y
17	61.8	5.4	625	13	BM628275	BM628275 170006874
18	60.8	5.3	698	13	BI574404	BI574404 RH23418.5
19	60.2	5.3	432	12	BE756667	BE756667 210881 MA
20	60.2	5.3	780	10	BE034697	BE034697 MLO3B10 M
21	59.4	5.2	492	12	BC366606	BC366606 HVSME1000
22	59.4	5.2	508	10	AM465736	AM465736 BP230020A
23	59.4	5.2	633	14	BO472076	BO472076 HV04F10F
24	59.2	5.2	679	12	BF627089	BF627089 HVSMEB000
25	59.2	5.2	728	9	AA942405	AA942405 LD26593.3
26	59	5.2	457	10	AM220586	AM220586 EST96971
27	59	5.2	528	10	AM220587	AM220587 EST96972
28	59	5.2	541	10	AM220588	AM220588 EST96973
29	58.4	5.1	648	9	AI388895	AI388895 GH19951.5
30	58	5.1	683	12	BG125422	BG125422 EST471068
31	57.8	5.1	520	13	BM429385	BM429385 IA20E01 B
32	57.8	5.1	561	12	BF618417	BF618417 HVSMEC000
33	57.6	5.0	685	13	BI353054	BI353054 CM22035.5
34	57.6	5.0	921	17	CNS078AH	AL433743 T7 end of
35	57.2	5.0	671	13	BI627668	BI627668 RH69047.5
36	57	5.0	467	12	BE756666	BE756666 210880 MA
37	56.8	5.0	994	17	CNS078X5	AL434559 T7 end of
38	56.4	4.9	653	13	BI624730	BI624730 RH64347.5
39	56.4	4.9	664	13	BI616334	BI616334 RH64249.5
40	56.2	4.9	672	9	AU001142	AU001142 AU001142
41	55.8	4.9	449	10	AW82490	AW82490 LG1_313-A
42	55.8	4.9	712	14	BO805492	BO805492 WHE3567_E
43	55.8	4.9	720	14	BO120882	BO120882 EST606458
44	55.6	4.9	488	10	BE215526	BE215526 HV_CEB000
45	55.6	4.9	617	10	AV939921	AV939921 AV939921

ALIGNMENTS

RESULT 1	CNS0778D	1057 bp	DNA	linear	GSS 08-JUN-2001
LOCUS	CNS0778D	T3 end of clone BB0A009H12 of library BB0A from strain CBS 4732			
DEFINITION	T3 end of clone BB0A009H12 of library BB0A from strain CBS 4732				
ACCESSION	AL432371				
VERSION	AL432371.1	GI:12215785			
KEYWORDS	GSS.				
SOURCE	Pichia angusta.				
ORGANISM	Pichia angusta				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.				
AUTHORS	1 (bases 1 to 1057) Souciet,J.L., Aigle,M., Attienave,F., Blandin,G., Boivin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozler-Kalogeropoulos,O., Potter,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNLT	FEMS Lett. 487 (1), 3-12 (2000)				
PUBMED	20584711				
REFERENCE	2 (bases 1 to 1057)				
AUTHORS	Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Attienave,F. and Dujon,B.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta				

[illegible]

Db	421	ACGCAATAGCCGTTCACAAAGCTATGCTGC CGGAAAACAAGAGATTCCACGACGGCTTGC	480
OY	800	ACTGCTCTGTTGTTCTCCTCAGACATTTACCACCTCAATTGTGCCACAGGACCTTTGGAA	859
Db	481	ACTGCTCTGGAATTCCTCAGACATTTCCACCACTCTATTGTGCCACAGGCGCTTTGGGAA	540
OY	860	TTCGCGCTCATATGTCGCCGTTTGCGGGAGACACCAATTTGATTCATGCACCAATGCTCTGA	919
Db	541	TTCGCGCTCAACGTGGCCATTGGGGAGACACCACTTTGATTCATGCACCAATGCTCTGA	600
OY	920	CTTACACGAGAATAATACGCTAACCGGGCTCCATGTGCTTACACCGTTCAGAGACTTCCAGAG	979
Db	601	CTTACCAAGAAGAAATATACGCTAACCGGGCTCCATGTGCTTACACCGTTCAGAGAG	660
OY	980	TTTGTCAGAGCGCTTGGAGATGTCATTAATCTTTGGACAAAGCGCCAGATGATTACAG	1039
Db	661	TTTGCGGGCGCTTGGAGATGTCATTAATCTTTGGACAAAGCGCCAGATGATTACAG	720
OY	1040	GCAAAGTCCACCTTAAGACGGAGTCGAGAGGGCTTTAACAAGCTGATTCGACACACAAG	1099
Db	721	GCAAAGTCCACCTTAAGACGGAGTCGAGAGGGCTTTAACAAGCTGATTCGACACACAAG	779
OY	1100	AGAACATGTCAAGATCCTG-GTGAGCGCCGAGACAGAGTTTCTTA	1142
Db	780	AAACCAATGTCAAGATCCTG-TGTGACTTCGACAGAGGTGTCTTA	823
RESULT 2	CNS07444/c	893 bp DNA linear GSS 07-JUL-2001	
LOCUS	CNS07444	clone BA0AB025B04 of library BA0AB from strain CLIB 210 of	
DEFINITION	Kluyveromyces lactis, genomic survey sequence.		
ACCESSION	AL428330		
VERSION	AL428330.1 GI:13432285		
KEYWORDS	GSS.		
SOURCE	Kluyveromyces lactis.		
ORGANISM	Kluyveromyces lactis		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.		
JOURNAL	1 (bases 1 to 893)		
MEDLINE	Soicler,J.L., Aigle,M., Artiguenave,F., Blandin,G.,		
PUBMED	Bouchier-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,		
REFERENCE	Duchateau-Nguyen,G., Lemaire,M., Marmelisse,R., Montrocher,R.,		
AUTHORS	Robert,C., Termier,M., Winkler,P. and Wesolowski-Louvel,M.		
TITLE	Genomic exploration of the hemiascomycetous yeasts : 11.		
JOURNAL	Kluyveromyces lactis		
MEDLINE	FEBS Lett. 487 (1), 66-70 (2000)		
PUBMED	20584721		
REFERENCE	11152886		
TITLE	3 (bases 1 to 893)		
JOURNAL	Direct SubMISSION		
MEDLINE	Submitted (08-Sep-2000) Genoscope - Centre National de Sequençage,		
PUBMED	2 rue Gaston Crepeux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :		
REFERENCE	segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
AUTHORS	This GSS is part of a random genomic sequencing program of thirteen		
TITLE	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces		
JOURNAL	exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,		
MEDLINE	Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces		
PUBMED	lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia		
REFERENCE	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,		
AUTHORS	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to		

Db 233 AGAGAAAAATCATGATTTTGTACACCTTAA 201

Db 945 AACTGAGAACGCGACGTTTCCAAAGTCAAGTTGGTATACAGTGGTGTGTAAGTGCATNGCT 886

Qy 308 CGCATTTGCTCCGACAGATGCGCTACAAAGACAGGTCGCCCCAAGACCTTTGGCTCTGTA 367

Db 885 GCACGTGTGAAGACCGCGCTAGGTGGCCCGACGCTCAAGACGAGAGGAGAGGCTCTGTSA 826

Qy 368 TGGCCTTGCCAGACCGATCTCCGAACTGCTGTGCTGGCTGACCTCTCTGGCTTGGGATG 427

Db 825 TTTCATGTCAAGACAAGAAACCTTACTGCTGCTGACCTGACCTTGGATTTCTGTGGGCTTGGGT 766

Qy 428 GTGCCAAGCGCGGTTTTTGGCCGAGTACGCTCTTTACGGTGAAGACCAACATGTCAGCTGC 487

Db 765 GCCAAATGTGGCGGTTTTTGGTGAAGAAATTTGCTTCACGACGCGTAATGTGCTCAAAATTC 706

Qy 488 CAGACTTCGATCCCGACGATATTGGACACTGTGGTTAGCCTATTTCCTGTGCTCGGATG 547

Db 705 CAACAACCTCTCCCTCGATGTGGCCGACCTTTGAGAACCCTATCTGCTCTTGGGACG 646

Qy 548 CTGTGTAAGCGGCTAGATATTCAGCTGTGTGAGACGCGCTGTGCTTGTGAGAGAGTCTTA 607

Db 645 CGGTCCGATTTTCGAACCTACACACAGCTGACACGCTTTGGTTTGGGTGCTGGCCCA 586

Qy 608 TCGGCGCTTGGCCACCATTTCTTGTCTGTGCAAGGCCATCAATCGCGGCAAAATTTGTGTTCCG 667

Db 585 TC -SSCTAGCGGCGAGTTCTTGTGCTGCACTGACAGAGCCATGTGTGCGGACCATTAATTTCTAG 527

Qy 668 AGCGCGCTTGATAGAAAGACAGTTTGCAAAGAAAGTGGGCGCTGGAAGTGTTCATCTT 727

Db 526 AACTGCGACACTAGAAAGCGCAAAAGCTGAGGCTTTGGTGTCTTCAAACTTCATCAATCAA 467

Qy 728 CTACA - - TGTGACAGCGCAAAATGCTGTTTCAAGGCTATGTGTCGCGAGACAGAGGAT 784

Db 466 CAGAACTAGTGATTAAGCTCATCGAAAGGCGTTCTGTAATTTGGCTCCAGGCGGTGAAGTT 407

Qy 785 TCCATGACAGCTTGGACTGCTCTGTGTTCTCTGACACTTACCAACCTCAATTTGCGCCA 844

Db 406 TCAACTTCTCTACGACGTGTGTGGTACACGAGACTTTTACGCGCGGCTGCAATAC 347

Qy 845 CGGACCTTTCGATGATCGCCCTCAATGTGGCGGTTTGGGAGACCCCAATTTGATTTCA 904

Db 346 TAACACTTAAGGTGTGCTGTCAACATCGGGGTTTGGGGTTCCTTAAGCTTAATGACTTCT 287

Qy 905 TGCCATGTCTGTGACTTAACAGAGAAATACGCTACCGGCTCATGTGCTAACCCGCA 964

Db 286 ACCCAATGATGTCTCAACATCAAGAGAAATTTGTTACAGGATCACTGTCTACACTACTC 227

Qy 965 AGGATTTCCAGAAATTTGTCAGAGCTTGAAGATGTGCTATATCTTTGGACAAAGCGC 1024

Db 226 AAGATTTTGAAGAGGTCTGACAGCAATGTGACAAAGGCAAAATTTACATCAACAAGTGA 167

Qy 1025 GCATAGATTAATACAGGCAAAAGTCCACCTTAAGAGACGAGTGTGAGAAAGGCTTTAAACAGC 1084

Db 166 AAACCATGATATCACAGTGTGCGCAAAAGTAATTTGAAGATGTTTTCAGAAAGGCAATTAAGATT 107

Qy 1085 TGATGAGACAGAGAGAAACATGTCAAGATCTGTGTCAGCGCGCAAGCA 1133

Db 106 TGATCAACAATTAAGAGAAAGAACATCAAAATTTTATGTAGCTCCGAACAA 58

RESULT 4
CNS06KBW 1066 bp DNA linear GSS 17-JUN-2001
LOCUS T7 end of clone AF00A004A07 of library AF00A from strain CBS 4311
DEFINITION of Saccharomyces servazzii, genomic survey sequence.
ACCESSION AL402690
VERSION AL402690.1 GI:12161979.
KEYWORDS GSS.
SOURCE Saccharomyces servazzii.
ORGANISM Saccharomyces servazzii.
REFERENCE
1 (bases 1 to 1066)
Saccharomycetales; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycotina; Saccharomycetes.

FEATURES	source
BASE COUNT	323 a 253 c 233 g 252 t 5 others
ORIGIN	
Query Match	17.1%; Score 195; DB 17; Length 1066;
Best Local Similarity	52.4%; Pzed. No. 1.1e-46;
Matches 476; Conservative	2; Mismatches 422; Indels 9; Gaps 2
misc_feature	<2..>967 /note="similar to Saccharomyces cerevisiae ORF YAL061w [FUM50 : similarity to alcohol/sorbitol dehydrogenase 1]" /evidence=not experimental
Y	234 CGGAACGGTGGTCAGGTTGGCTCTGGTGTCACAAATGGTGAACCTGGTGACAGAGTGC 293
Db	1 CGGTCTCTGTGANAAGATCGTCCCGGTGTCAAAATTCAAAGTCGGGACACAGTCTC 60
Y	294 AGTTAAGCTACGTGCGATTGCTCCGACAGATCGCGCTCAAGAGACAGGTCGCCAAGA 353
Db	61 TGTAAACCTACAGTACATGTAGAGACAGAGTACAGATGGCAGACGCTCCAAAGCAAA 120
Y	354 CCTGGGCTCTGTATGGCTTCCGACAGAGGAGATCTCGAAGCTGTGGCGTCTGAGCTT 413
Db	121 CAACCAAAATGCCAGATGTAAAAAGGSAATGTACATTACTTGCTTGGCTACTTGGGGCT 180
Y	414 CTGGCGTTTGGTGATGTCACAGCGCGGTTTGGCGGATAGCTGTTACGGTGAGACCA 473
Db	181 AATCGTAAACGGGGTCCAAAGTGGTGGTTTAGTGAACGATGGTCAATCAAGATACACA 240
Y	474 CATGTCAAGCTCCAGACTGATTCGCCAGCATATTGGAGCAGCTGGTGAACCTATTTC 533
Db	241 CTGTTCACAAAGATCCCAACCATATCCACATAGAGTGGCTTTTATCCAACTATTTC 300

OY	534	TTGTTCCTGGCATGCTGTTGAAGACGGGCTTGATTCACAGCCCTGGTTCACAGCGCCCTGGTCT	593
Db	301	TTCTCACTTGGCATGCTGCTGGTGTCTTCCAGATACAGACAAAGACGGGTCCCTCTTGAATCCT	360
OY	594	TGGAGAGAGTCCCTATTCGSCCTTGGCCACATCTTGGCTGTGCAAGGCCATCATCGGGCAA	653
Db	361	CGGGGGTGGGCAACATCGGCCCTAACCAATCTTGGCTTGAATGAGACACGGGTGTACCGA	420
OY	654	AATTTGTGTTTCCGACCGGCGCTTGATCGAAGACAGTTTGCACAAAGAACTGGGCGCTGA	713
Db	421	TATGCTCGTTTAGAACCAGCCAAATATGAGAAAGATCTAAGCAGAAARATGGATGGCCG	480
OY	714	AGTGTTCGATCCTTCTA-----CATGTACGACGCAAAATGCTGTTCTCAAGGCTATGTT	767
Db	481	TGTCTTCGATCCACAGCAAACTGTGCAAGAGAAAGAACCAAGATCCTTAGATCAATGGC	540
OY	768	GCCGGAGAACGAGGGATTCATGACAGCCTCTGCATCCTCTGGTGTTCCTCAGCATTTAC	827
Db	541	TCCAGGCGGTGACGGGTTGACTACTCGTTCGATCTGTCAAGTGTTCAGGTGTTCCAGCCACTTAG	600
OY	828	CACCCATATTTGTGCGCACGGGACCTTCTGGAATTCGCCCTCAATGTGGCGTTTGGC---G	884
Db	601	AGCTTCATTCGAAATCTTAACATTTTCAGAGGTGCCGTGTCCATGTGCGCATGTGGGACCG	660
OY	885	AGACCACCAATTTGATTCATGCAATGTGCTGTGAATTCACAGAGAAATACGCTACCGG	944
Db	661	TGGGAACATCTGCAATTTCTCCCAATGATATCACAAAACAAGAAACATCTACACCG	720
OY	945	GTCGATGTGCTAACACCGTCAAGGACTTCCAGGAATGTGTCAAGGCTTGGAAAGTGTCT	1004
Db	721	GTCGATGTGTTACACCTTATCACGATTTGCAGAAAGTCATCGACGCTTGGAAAAGGGCTT	780
OY	1005	CATATCTTTGGACAAGCGCGCAAGATGATTCAGGCAAGTCCACTTAAGAGACGAGT	1064
Db	781	GATCGATCCAGAAAGCGACGACCATGATTAAGTCCAAAGGTCCCTTGGAAAAGACTTT	840
OY	1065	CGAGAAAGGCTTTAAACAGCTGATGAGAGACAGAGAAACATATGCAATCTCTGTGAC	1124
Db	841	TGAGACGCGTATCATGATTAATTAATACACACACAAAGAAACCTATATCAATAGTGTGATCAC	900
OY	1125	GCCGAAACGA 1133	
Db	901	TTCCAACAA 909	

RESULT 5	CNS06193	976 bp	DNA	linear	GSS 17-JUN-2001
LOCUS					
DEFINITION	77 end of clone A70AA011E1 of library A70A from strain CBS 4311 of <i>Saccharomyces servazzii</i> , genomic survey sequence.				
ACCESSION	AL403885				
VERSION	ALA03885.1	GI:12164699			
KEYWORDS	GSS.				
SOURCE	<i>Saccharomyces servazzii</i> .				
ORGANISM	<i>Saccharomyces servazzii</i> .				
REFERENCE	Eukaryota; Fungi; Ascomycota; <i>Saccharomycotina</i> ; <i>Saccharomycetes</i> ; <i>Saccharomycetales</i> ; <i>Saccharomycetaceae</i> ; <i>Saccharomyces</i> .				
AUTHORS	1 (bases 1 to 976) Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., deMontigny,J., Dujon,B., Durren,P., Lepingle,A., Lorente,B., Malpertuy,A., Neuvéglise,C., Ozler-Kalogreopoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett.	487 (1),	3-12	(2000)	
MEDLINE	20584711				
PUBMED	11152876				
REFERENCE	2 (bases 1 to 976) Casaregola,S., Lepingle,A., Bon,E., Neuvéglise,C., Nguyen,H., Artiguenave,F., Wincker,P. and Galliardin,C.				

TITLE	Genomic exploration of the hemiascomycetous yeasts: 7.
JOURNAL	Saccharomyces servazzii
MEDLINE	PMID: 11152882
PUBMED	20584717
REFERENCE	3 (bases 1 to 976)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
COMMENT	This GSS is part of a random genomic sequencing program of thirtee yeast species: Saccharomyces bayanus var. varium, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces fragilis, Kluyveromyces marxianus var. marxianus, Pichia candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES	<p>Location/Qualifiers</p> <p>1..976</p> <p>/organism="Saccharomyces servazzii"</p> <p>/strain="CBS 4311"</p> <p>/db_xref="taxon:27293"</p> <p>/clone="AT0AA011E11"</p> <p>/clone_lib="AT0AA"</p> <p>/note="end : 17"</p> <p><1..>972</p> <p>/note="similar to Saccharomyces cerevisiae ORF YAL061w [FUN80 ; similarity to alcohol/sorbitol dehydrogenase]"</p> <p>/evidence="not_experimental"</p> <p><7..>969</p> <p>/note="similar to Saccharomyces cerevisiae ORF YAL060w [FUN89 ; similarity to alcohol/sorbitol dehydrogenase]"</p> <p>/evidence="not_experimental"</p>
BASE COUNT	291 a 151 c 240 g 270 t 24 others
ORIGIN	
Query Match	16.1%; Score 184; DB 17; Length 976;
Best Local Similarity	50.8%; Pred. NO. 2e-43;
Matches 479; Conservative 13; Mismatches 443; Indels 7; Gaps 3	
QY 180	CAAGCAACAAGATTGGGATACGACTTCCTCTCTGCTCGACATGAATTTAGCGGAAC 239
Db 3	CAGACATACGCTACTGAAATRAACTTCCCGAGCTCTTGACACGAGTACAGTGTTA 62
QY 240	GATGTCGAGGTTGGCTCTGCTGTGCACAAAGTGTGAACCTGGTGACAGAGTCGAGTTGA 299
Db 63	TGTGAACGMAFTTGGACGACGCGTGACGTCATTAATMAAGTAGGAGCAGAGTGTGTCGA 122
QY 300	AGCTACGTCGATTCCTGCCACAGATTCGCGCTACAAAGACACGTTGCCCAAGCCTTGG 359
Db 123	AGCAAGTCTAGTWTGTTGATTAACATCTGTGGCCCAATGCCAAMWTTGTAACAMACC 182
QY 360	GCTCTGATGACCTGCCACGAGGAGATTCGCCAAGTCGTCGCTGCGCTGAGCTTCTGGG 419
Db 183	TATATGTCATCTGTTAATGAAGGAMACGAATCTGTTGTGAATATGCTGTTTACGTGC 242
QY 420	TTTGGGTGTCACGACGGCGGTTTTGCGGAGTAGCTGCTTTACGGTGAGGACACATGGT 479
Db 243	TTTGGGTGTAATTTCTGAGGAGGTTTGCAGACACGATGGAACCATTTGAACATCATGTGT 302
QY 480	CAAGCTGCACATTCGATTTCCGACAGATTTGAGACACTGGTTGAGCCTATTTCGTTGC 539
Db 303	TAAAMWACCTGATTTCTTACCTATGAGATGTTGGTCTTATGTGAGCATTGCTGTTGC 362
QY 540	CTGACATGCTGTTAGACGCGCTAGATTTCCACCTGGGTGAGACGGCCTGGTTCTGGAGG 599
Db 363	ATGCGATGCTTCGTGTGAGCMAATTTTACCCCTGTGAAGACTGCTTTAAMWCTTGTTGC 422
QY 600	AGGTCTTATGGGCTTGCACCACTATTCTTGCTGTGCAAGGCATCATGCGGCGCAAAATGT 659

```

Db 423 AGTCCAAATGGCGMAGCCATGACCTGGTGTAAAGGCCAAARGTGCTAAAGATTTGT 482
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 660 GTGTTCCGAGCCGCTTGATCAGAAAGACATTTGCAAGAACTGGCGCTGAAGTGT 719
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 WGTAGTGAACAGGATCATATCAAGAAATTTGGCTGAAGAACATCATGTGAAACATT 542
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 720 CGATCTTTTGA--CATGTGACAGAGCAATGCTGTTCAGAGGCTATGTGCGGAGAA 776
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 543 TGMCTCTCAAGACATGAGATTAATTCMAATTAACCGAATAGCTTGCAACAGAAAGTAA 602
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 777 CGAGGATTCATGACGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 836
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 603 TAAAGGCTTTGATTTTGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 662
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 837 TGTGCGCAGCGGACCTTTCGAAATGCGCGCTCAATGTGGCGGTTGGGAGACCCCAAT 896
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 663 TGTGCTATMCATTAATAGAGTACTTATTTGTAACGTTCGCAATTTGGGCAAGGATTTAA 722
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 897 TGGATTCATGCCATGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 723 T--TTTAAACCCAAATGATTAATTAATTAACAGAAAGAAATTTGACTGTTCTATTGTTGA 779
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 957 CACCTCAAGAGACTTCAGAAATGTCTCAAGGCTTGAAGATGCTCATATCTTTGGA 1016
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 780 TACAGTAGAAGATTTTAAAGCAAGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 839
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1017 CAAAGCGCGCAGATTAATTAACAGCAAGTCCACTTAAAGAGCGAGTGCAGAGGCTTT 1076
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 840 AGAGTGTGAACATTTGATTCACCGGTAGACAAAATTTGAA-GATGTTGGGAAAAGGGGTT 898
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1077 TAAACAGCTGATGACAGCAAGGAAACATGTCAGATCCT 1118
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 899 CTTGAGTGTGATTAACCATTAAGATTAACATCAATCAAGTCT 940
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
CNS06XOA/ 994 bp DNA linear GSS 06-JUL-2001
LOCUS T3 end of clone AY0AA003A03 of library AY0AA from strain CBS 6340
DEFINITION of Kluyveromyces thermotolerans, genomic survey sequence.
ACCESSION AL420056
VERSION AL420056.1 GI:12203234
KEYWORDS GSS.
SOURCE Kluyveromyces thermotolerans.
ORGANISM Kluyveromyces thermotolerans.
REFERENCE 1 (bases 1 to 994)
AUTHORS Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
1 (bases 1 to 994)
Bouclet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Porter,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 994)
AUTHORS Malpertuy,A., Llorente,B., Blandin,G., Artiguenave,F., Wincker,P.
and Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 10.
Kluyveromyces thermotolerans
JOURNAL FEBS Lett. 487 (1), 61-65 (2000)
MEDLINE 20584720
PUBMED 11152885
REFERENCE 3 (bases 1 to 994)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

```

```

COMMENT
sequel:genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia
lactis var. lactis, Kluyveromyces fragilis var. fragilis, Candida
tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
    Location/Qualifiers
        1..994
        /organism="Kluyveromyces thermotolerans"
        /strain="CBS 6340"
        /db_xref="taxon:4916"
        /clone="AY0AA003A03"
        /clone_11b="AY0AA"
        /note="end : T3"
        complement(6..>671)
        /note="similar to Saccharomyces cerevisiae ORF YAL060w [
        FUN49 : similar to alcohol/sorbitol dehydrogenase ]"
        /evidence=not_experimental
        complement(<66..>671)
        /note="similar to Saccharomyces cerevisiae ORF YAL061w [
        FUN50 : similar to alcohol/sorbitol dehydrogenase ]"
        /evidence=not_experimental

misc_feature
    complement(66..>671)
    /note="similar to Saccharomyces cerevisiae ORF YAL061w [
    FUN50 : similar to alcohol/sorbitol dehydrogenase ]"
    /evidence=not_experimental

misc_feature
    complement(66..>671)
    /note="similar to Saccharomyces cerevisiae ORF YAL061w [
    FUN50 : similar to alcohol/sorbitol dehydrogenase ]"
    /evidence=not_experimental

BASE COUNT 243 a 263 c 242 g 244 t 2 others
ORIGIN
Query Match 13.9%; Score 159.2; DB 17; Length 994;
Best Local Similarity 54.4%; Pred. No. 5.6e-36;
Matches 365; Conservative 0; Mismatches 299; Indels 7; Gaps 2;

Oy 1 AAGAAAGTTTACTTTATAGAGTAAACGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 ATGAGAGCTTGAACATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 612
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 61 CCGAGATCAAGAAATCCCAACAGATGCAAGATCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 CCACAGGTTTCAATGAGATGAGTGGAATTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 552
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 121 ACGGACTTAAAGAAATTCATATCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 551 TCCGATCTCCAGCA-----ATATCTTGAGGCCCTTATTTCTTCCAGCTAATGGAAG 498
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 181 AAGGACAAGATTTCCGGATACGAACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 ACACACGCTATGAGCGCCCAAGGCTTCCCAAGGCGCATGGCCATGAGATGCTGGCAATT 438
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 241 GTGGTGGAGGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 437 GTGTCAAAAGTAGGAAGCCACGCTTCCAAAGGCTTCAAGGCTTGAATGATGATGATGATGATGATGATGAT 378
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 301 GCTACGCTGCGATTTGCTCCGACAGATGCGCTCAAGAGACAGCGATGCGCCCAACACCTTGGG 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 GCATGCTGACGCTGTGAAGCCCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 361 CTCTGTATGAGCCCTGCGACAGCGGATCTCCGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 317 CTGTGATTCATGCAAGAGCAAGAAACCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 421 TTGGGTGGTCCAGCGCGGCTTTTGGCGAGTACGTGCTTACGGTGAAGACCAACATGTC 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 257 CTTGCGGTGCGCAAGTGGCGGCTTTTGTGAAAAAATTTGCTCTAGCAGAGGCTAATGTGCTG 198
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 481 AAGCTGCAAGACTCGATTCGCCGAGATTTGAGCACTGTGAGGCTATTCTGTGTC 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 197 AAAATTCGAACAGTCTTCCCTCGATGTGGCGCACTGTGAACCTAATCTCTCTCTCT 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 541 TGGCATGCTGTGAACGCGCTAGATTCAGAGCTGTGTGAACGCGGCTGCTTCTTGAGAGA 600
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137 TGGCAGCGGCTCGCATTTGCAAGCTACAAACGCTGACAGCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 78
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

BASE COUNT	252 a	160 c	232 g	268 t	11 others
Query Match	13.0%: Score 148.8; DB 17; Length 923;				
Best Local Similarity	52.6%: Pred. No. 6.9e-13;				
Matches 367; Conservative	7; Mismatches 306; Indels 18; Gaps				
9	TTTACTTATATACGCTCAACACATATTCGCTACTCGCAACGGTTCCTGAAACCGGAGAT	68			
Db	145 TTTCGTTATTTTGGAAACAGAGATATTCCTTTTCMCACATGATCTACCTGAAACCGATAT	204			
QY	69 CAAAGATCCCAACGATGTCAAGATCAAGATGACACTATTTGGAATATGTGGACACGACTT	128			
Db	205 TGAAGAACCGGATGATTTTGTATCCATATATGCTGCTGCTGCTGCTGCTGCTGCTGCT	264			
QY	129 GAAAGATTTCAATATTTCTGGAGTCTCTGTTTTTTTTTCCCTAAACAAGCACCAGCAACAA	188			
Db	265 ACATGAACCTCTGAC-----GGACCAAACTCTTCCCAAAAGATGTGAACAACATGA	318			
QY	189 GATTTCGGGATPACGAATCTCTCTCTGCTGACATGAAATTTACGGAAACGCTGTGCA	248			
Db	319 DATTAAGTGTAAGGGATTTGCTCTCAACCATGGCCACAGATGCAAGCGCTGTTTCA	378			
QY	249 GATTGCTCTGCTGCTCAACAAGTGTGAACCTGTGTGACAGAGTGCAGATTGAGCTACGTC	308			
Db	379 AGTCGGCTCTAGTGTACATAATTTCCACACAGATACCATTTGTTGTGGAGCTACAGG	438			
QY	309 GCATTGCTCCGACAGATGCCGCTACAAAGACAGCGGCCCAAGCTTTGGGCTCTGAT	368			
Db	439 TACATCTGACGAGACAGAAAGATATATATCTAAG-----GTGADACGTGCA	486			
QY	369 GGGCTCCACAGACGAGATCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	428			
Db	487 AGTTGTAAAGGGGATGACACAAATTTGTGTGATTTGGGTTATGCGGATGTGGTGT	546			
QY	429 TGCACAGCGCGGTTTTTGCAGATGATGCTGTTAGCTGAGAGACACATGCTGCAAGCTGCC	488			
Db	547 ACAATCAGGTGATTTTGGTGAAGAAAGTTGTTATCGGTGAGAACACATGTTTCAAGATACC	606			
QY	489 ACACGCGATATCCCGACGATATTTGAGACGACGCTGTGAGCCTATTTCTGTTCCTGGCATGC	548			
Db	607 ATCATGATATCCCATGTTGTTGGCGACGATATTTACGCTTTTACAGCTTTGCTGGCATGC	666			
QY	549 TGTTAAGCGCGCTAGATTCACAGCGCTGACACAGCGCCCTGTTTGTGGAGAGAGCTCTAT	608			
Db	667 TGTGAATTTGGGAGCTGTGAAGAAAGGTCATCTGCAATGATTTCTCGTGTGTGCTGAT	726			
QY	609 CGGCTTCCGACATCTTCTGCTGCTGCAAGGGCATCATGCGGGCAAAATTTGTGTTCCGA	668			
Db	727 TGGGTTGGGCAATTTAAAGGCTTTGCAACACTTTGGGTGACATGATATCGTGTTCGGA	786			
QY	669 GCCGCCCTTGATCAGAAAGACAGTTGGCAAAAGAACCTG 706				
Db	787 GCTGCCAAATTTAGAGCTATCTAGCTGAAAAATGGG 824				
RESULT 8					
LOCUS					
CNS07443					
DEFINITION	CNS07443 908 bp DNA linear GSS 07-JUL-2001				
ACCESSION	clone BA0AB025B04 of library BA0AB from strain CLB 210 of				
VERSION	AL428329				
KEYWORDS	AL428329.1 GI:13449456				
SOURCE	GSS.				
ORGANISM	Kluyveromyces lactis.				
REFERENCE	Kluyveromyces lactis.				
AUTHORS	Eukariyota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; Saccharomycetaceae; Kluyveromyces.				
	1 (bases 1 to 908)				
	Soullet, J. L., Aigle, M., Artiguenave, F., Blandin, G.,				
	Boulotin, F., Kuhn, M., Bon, E., Brotier, P., Casaregola, S.,				
	de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, R.				

	TITLE	Sautin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Winkler,P. and Weissensbach,J.
JOURNAL MEDLINE PUBLISHED		Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
REFERENCE AUTHORS		FEBS Lett. 487 (1), 3-12 (2000)
TITLE		
JOURNAL MEDLINE PUBLISHED		2 (bases 1 to 908)
REFERENCE AUTHORS		Biolotin-Fukuhara,M., Toffano-Nioche,C., Artiguenave,F., Duchateau-Nguyen,G., Lemaire,M., Marneise,R., Montrocher,R., Robert,C., Templier,M., Winkler,P. and Wesolowski-Louvel,M.
TITLE		Genomic exploration of the hemiascomycetous yeasts: 11.
JOURNAL MEDLINE PUBLISHED		Kluveromyces lactis
REFERENCE AUTHORS		FEBS Lett. 487 (1), 66-70 (2000)
TITLE		
JOURNAL MEDLINE PUBLISHED		3 (bases 1 to 908)
REFERENCE AUTHORS		11152886
TITLE		
JOURNAL MEDLINE PUBLISHED		Genoscope.
REFERENCE AUTHORS		Direct Submission
TITLE		Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage.
JOURNAL MEDLINE PUBLISHED		2 rue Gaston Creteilux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segre@genoscope.cns.fr Web : www.genoscope.cns.fr)
REFERENCE AUTHORS		This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces servazii, Zygosaccharomyces rouxii, exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Kluyveromyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii Var. hansenii, Pichia sorbitophila, Candida tropicalis and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
COMMENT		
FEATURES		
source		Location/Qualifiers 1..908 /organism="Kluveromyces lactis" /strain="CLIB 210" /variety="lactis" /db_xref="taxon:28985" /clone="BA0AB025B04" /clone.lib="BA0AB" 51..>571 /note="Similar to Saccharomyces cerevisiae ORF YAL060w [FUNd9 ; similarity to alcohol/sorbitol dehydrogenase] similar to Saccharomyces cerevisiae ORF YAL061w [FUNd50 ; similarity to alcohol/sorbitol denhydrogenase] " evidence-not-experimental
misc-feature		
BASE COUNT		298 a 160 c 191 g 259 t
ORIGIN		
Query Match		10.5%; Score 120.4; DB 17; Length 908;
Best Local Similarity		53.3%; Pred. No. 2.2e-24;
Matches		304; Conservative 0; Mismatches 256; Indels 10; Gaps 2
OY	575	GTCAGACGGCCCGGCTTGTGGAGGAGGCTATCGGCCCTTGCCACCATTCTTGCTCTGC 634
DB	1	GTCAACTCATGATGTTCTCGGTGGGTCGCCGATGGGTTGGCATTATATGCGCTTGG 60
OY	635	AAGGCATCATCGCGGCAAAATGTGTGTCGACGCGCTTGATCAGACAAGACACTTGG 694
DB	61	AMCACTTGGGTGCTACTGATATGCTGGTGTGGAGCGCTGCCAANAATTAGACGCGATCTAG 120
OY	695	CNANGAACACGGGCGCTGAAGTGTTCGATCCCTCTACATGTGACGACGCAATGCTGTT 754
DB	121	CTGAAAAAATGGGTGCTATTTCCAAACCTTAGATTTCAAGACCATAAGAGAAAACA 180
OY	755	TCAAAG-----CTATGGTCCGGAAGAAGAGGATTCATGCGACCTTGACTGCT 805
DB	181	TCAAAACTTATGCTTCATTTATCTCCAATGGTGAAGAGTAAGATTTTCAATTGATTTGTT 240
OY	806	CTGGTGTTCCTCANACATCACCAACCTCAATATGTCGCGACGAGGACTTCTTGATTCGCG 865
DB	241	CGGATTTTCTGTCACCTTTCGATTCATCATCAATGCTTCAACTTCAAGAGGAACTGCA 300

OY 866 TCATATGTGGCGTTTGGGGAGACACCACCAATTGGATTTCATGCCAATGCTCTGACTTACC 925
 Db 301 TCATATGTGGCGTTTGGGGAGAGCAGCTCTCTTGAAGCATTAATTCATGAATGATTTTAACATCAC 360
 OY 926 AGGAGAAATACGCTACCGGCGTCATGTGCTACACCGCCGCAAGCATTCACGAGATTGTCA 985
 Db 361 AGGAAAAAATATACAGAGGCTCATATGTTTACACATTTGAGAGATTTTGAGAGAGCTGTG 420
 OY 986 AGGCGCTTGAAGATGATCTCATATCTTTTGGACAAAGCCGCAAGATGATTACGGCAAG 1045
 Db 421 ATTCCTTCACCGA-AGTAAACAGATTTCAAAAGAGCGCTTCATGATCAGCGGAAAG 479
 OY 1046 TCCACCTTAAGACCGAGTGAGAGAGGCTTTAAACAGCTGATCGAGCAACGAGAGACA 1105
 Db 480 TTTCTCTGAGAAAGGATGAGATGAGACGCGATTCATGATGATTTGATCAATGACAAAGAGAA 539
 OY 1106 ATGTCAGATCTCTGCTGACGCCGAGCAGG 1135
 Db 540 CCATCAAGATCTCTACTTACTCTCAAGAGAG 569

RESULT 9
 CENS070R5/C
 LOCUS
 DEFINITION T3 end of clone AZ0A011F06 of library AZ0A from strain CBS 712 of Kluyveromyces marxianus, genomic survey sequence.
 ACCESSION AL423975
 VERSION AL423975.1 GI:12207169
 KEYWORDS GSS.
 SOURCE Kluyveromyces marxianus.
 ORGANISM Kluyveromyces marxianus
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE
 AUTHORS 1 (bases 1 to 924)
 Soulier,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durans,P., Lepingle,A., Florente,B., Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potlier,S., Saurin,P., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissensbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

TITLE
 JOURNAL JOURNAL
 MEDLINE 20584711
 PUBMED 11152876
 PUBMED 20584722
 JOURNAL 11152887

COMMENT
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Crepieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermocolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hanseni var. hanseni, Pichia stipitophila Candida tropicalis and both extremities were sequenced. See keywords for description of this insert.
 The other extremity of this insert.
 Location/Qualifiers
 1..924
 /organism="Kluyveromyces marxianus"
 /strain="CBS 712"

/varietal="marxianus"
 /db_xref="taxon:4911"
 /clone="A20A011F06"
 /clone_lib="A20A"
 /note="end : T3"
 complement(<1..>456)
 /note="similar to Saccharomyces cerevisiae ORF YAL060W [FUN49 ; similarity to alcohol/sorbitol dehydrogenase] similar to Saccharomyces cerevisiae ORF YAL061W [FUN50 ; similarity to alcohol/sorbitol dehydrogenase]"
 /evidence=not_experimental
 BASE COUNT 290 a 239 c 139 g 254 t 2 others
 ORIGIN

Query Match 9.4%: Score 107.4; DB 17; Length 924;
 Best Local Similarity 53.8%; Pred. No. 1.7e-20;
 Matches 247; Conservative 0; Mismatches 206; Indels 6; Gaps 1;

QY 1 ATGAAAGTTTACTTATTAAGGTGATCAAGATATTCGCTACTCCGAAAGCTTCCGAA 60
 DB 456 ATGCGTCCGTTAGCGTATTTCCGTAAGAAAGATTAATTAATTAATTAATTAAGAG 397
 QY 61 CCGGAGATCAAGATCCCAAGATGCAAGATCAAGATCAAGATCAAGATCAAGATCAAG 120
 DB 396 CCGGAGATGCGGTGCGAGATGCAAGATTAAGATTAAGATTAAGATTAAGATTAAG 337
 QY 121 ACCGACTTGAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 180
 DB 336 AGTACCTTCATCAATACCTAGAT-----GTCCTAATTTCTTCCCGCAAGATGGAAG 283
 QY 181 AAGGACAAATTTGGGATACGAATCTCTCTCTCTGATGATGATGATGATGATGATG 240
 DB 282 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 223
 QY 241 GTGCTGAGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 DB 222 GTGCTGAGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163
 QY 301 GGTAGTCCGATTTCTCCGACAGATGCGCTGCAAGAGACGCGTCCGACGACCTTGG 360
 DB 162 GCTAGTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 103
 QY 361 CTCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 102 AATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 43
 QY 421 TTGGGTGCTGCGGCGGCTTTTGGCGAGTACGCTT 459
 DB 42 CTGGGTGCTGCGGCGGCTTTTGGCGAGAAAGTTGT 4

RESULT 10
 CDS07600/c 1045 bp DNA linear GSS 08-JUL-2001
 LOCUS T7 end of clone BB0A005F01 of library BB0A from strain CBS 4732
 DEFINITION of *Pichia angusta*, genomic survey sequence.
 ACCESSION AL431664
 VERSION AL431664.1 GI:12215078
 KEYWORDS GSS.
 SOURCE *Pichia angusta*.
 ORGANISM *Pichia angusta*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; *Pichia*.
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durieux,P., Lepoint,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogiropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Mioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584721
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 1045)
 AUTHORS Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.,
 and Dujon,B.
 TITLE Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia*
angusta
 JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
 MEDLINE 20584723
 PUBMED 11152888
 REFERENCE 3 (bases 1 to 1045)
 AUTHORS
 TITLE Genoscope.
 JOURNAL Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*
exiguus, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 the keywords for description of this sequence and for the sequence of
 the other extremity of this insert.
 location/Qualifiers
 1..1045
 /organism="Pichia angusta"
 /strain="CBS 4732"
 /db_xref="taxon:4905"
 /clone="BB0A005F01"
 /clone_lib="BB0A"
 /note="end : T7"
 complement(<876..>977)
 /note="similar to Saccharomyces cerevisiae ORF YAL060W [FUN49 ; similarity to alcohol/sorbitol dehydrogenase]"
 /evidence=not_experimental

BASE COUNT 285 a 248 c 227 g 280 t 5 others
 ORIGIN

Query Match 8.8%: Score 100.2; DB 17; Length 1045;
 Best Local Similarity 83.0%; Pred. No. 2.8e-18;
 Matches 137; Conservative 1; Mismatches 24; Indels 3; Gaps 2;

QY 978 AGTTTCAAGCCTTGGAGATGCTCATATCTTTGACAAAGCGCGCAAGATGATAC 1037
 DB 1029 AGAAGTTCGGGCTTGGAGATGCTCATATCTTTG--ACAAGCGCGCAAGATATTAC 972
 QY 1038 AGCAAAAGTCCACCTAAAGAGCGAGTCCGAAAGGCTTTAAACAGCTGATCGACACA 1097
 DB 971 A-GCAAAAGTCCACCTAAAGAGCGAGTCCGAAAGGCTTTAAACAGCTGATCGACACA 913
 QY 1098 GGAACCAATGTCAGATCCTGCTGACCGCAAGAGGTTTCTTA 1142
 DB 912 GGAACCAATGTCAGATCCTGCTGACCGCAAGAGGTTTCTTA 868

RESULT 11
 CDS0619E 982 bp DNA linear GSS 17-JUN-2001
 LOCUS T7 end of clone AT0A011F10 of library AT0A from strain CBS 4311
 DEFINITION of *Saccharomyces servazii*, genomic survey sequence.
 ACCESSION AL403896
 VERSION AL403896.1 GI:12164726
 KEYWORDS GSS.
 SOURCE *Saccharomyces servazii*.
 ORGANISM *Saccharomyces servazii*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durieux,P., Lepoint,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogiropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Mioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 08:04:46 ; Search time 100 Seconds

(without alignments)
3505.315 Million cell updates/sec

Title: US-10-020-674-1

Sequence: 1143

Sequence: 1 atgaaggttacttactta.....cgccgaagaggttctctaa 1143

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6C.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/6D.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57.4	5.0	933	4	US-09-347-803-21
2	49.4	4.3	1065	4	US-09-134-001C-537
3	49	4.3	1610	4	US-09-347-803-19
4	47.2	4.1	2774	3	US-08-466-548B-1
5	47.2	4.1	2774	3	US-07-998-226F-1
6	47.2	4.1	2774	3	PCT-US93-12560-1
7	45	3.9	1393	3	US-08-174-467-18
8	45	3.9	1393	3	US-08-452-071-17
9	43.4	3.8	1419	3	US-08-174-467-18
10	43.4	3.8	1419	3	US-08-452-071-17
11	41.8	3.7	1154	4	US-09-347-803-15
12	40.6	3.6	1283	1	US-08-174-467-19
13	40.6	3.6	1283	1	US-08-452-071-19
14	39.8	3.5	30001	2	US-08-125-468-1
15	39.8	3.5	30001	2	US-08-474-933-1
16	35.8	3.1	6877	1	US-08-347-340-1
17	35.6	3.1	1188	4	US-08-860-656B-9
18	35.6	3.1	1481	4	US-08-860-656B-1
19	35.4	3.1	1377	1	US-08-174-467-20
20	35.4	3.1	1377	1	US-08-452-071-20
21	35.2	3.1	2061	4	US-09-800-960-1
22	33.2	2.9	2318	3	US-08-826-964-1
23	33	2.9	535	2	US-08-713-000-1
24	33	2.9	535	2	US-08-975-316-1
25	33	2.9	535	2	US-09-211-710-1
26	33	2.9	535	4	US-09-615-192A-1
27	33	2.9	578	4	US-09-615-192A-112

28	33	2.9	592	2	US-08-975-316-30	Sequence 30, Appl
29	33	2.9	592	4	US-09-615-192A-30	Sequence 30, Appl
30	33	2.9	1087	4	US-09-372-422A-29	Sequence 29, Appl
31	32	2.9	910	4	US-08-456-200B-16	Sequence 16, Appl
32	32	2.8	1035	1	US-08-891-254-8	Sequence 8, Appl
33	32	2.8	1035	2	US-08-891-254-8	Sequence 8, Appl
34	32	2.8	1035	2	US-08-891-254-8	Sequence 8, Appl
35	32	2.8	1035	2	US-09-030-270A-8	Sequence 8, Appl
36	32	2.8	1035	4	US-08-984-207-8	Sequence 8, Appl
37	32	2.8	1035	5	US-09-013-387-8	Sequence 8, Appl
38	32	2.8	1035	5	PCT-US96-08819-8	Sequence 8, Appl
39	32	2.8	2277	1	US-08-676-967-2	Sequence 2, Appl
40	32	2.8	2277	1	US-08-676-967-2	Sequence 2, Appl
41	32	2.8	2277	2	US-08-676-967-2	Sequence 2, Appl
42	32	2.8	7791	2	US-09-098-487-2	Sequence 2, Appl
43	32	2.8	7791	3	US-08-149-097D-23	Sequence 2, Appl
44	32	2.8	7791	3	US-08-949-386-23	Sequence 2, Appl
45	32	2.8	7791	4	US-08-984-709A-23	Sequence 23, Appl
					US-08-450-272-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-347-803-21

Sequence 21, Application US/09347803

Patent No. 6274379

GENERAL INFORMATION:

APPLICANT: Famodu, Layo O.

APPLICANT: Hitz, Bill

APPLICANT: Kinney, Tony

APPLICANT: Orozco, Buddy

TITLE OR INVENTION: Plant Sorbitol Biosynthetic Enzymes

FILE REFERENCE: BB-1176

CURRENT APPLICATION NUMBER: US/09/347,803

CURRENT FILING DATE: 1999-07-02

EARLIER APPLICATION NUMBER: 60/092,952

EARLIER FILING DATE: July 15, 1998

NUMBER OF SEQ ID NOS: 27

SOFTWARE: Microsoft Office 97

SEQ ID NO 21

LENGTH: 933

TYPE: DNA

ORGANISM: Trifolium aestivum

US-09-347-803-21

Query Match

Best Local Similarity 47.4%; Score 57.4; DB 4; Length 933;

Matches 172; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY	359	GGCTCTGTATGGCTCCAGAGCGGATCTCCGACTGCTGCTGCTGAGCTTTCGCG	418
DB	309	GGCGTTCGAGGCTCCAGAGCGGCGGCTTACACCTCTGACGACATGAATTCCTCG	368
QY	419	GTTTGGGTGCTCCAGAGCGGCTTTCGAGTACGCTGTTTACGCTGAGGACCATGCG	478
DB	369	CCACCCACCTTACATGATGATCTTCCGACACATTTGTGATCCAGTACCTGCTCT	428
QY	479	TCAGTCTCAGACTCGATTCCTCCAGCATTTTGGAGCATCTGTTAGCTTATTTCTGTG	538
DB	429	TCAGTCTCAGACTCGATTCCTCCAGCATTTTGGAGCATCTGTTAGCTTATTTCTGTG	538
QY	539	CCTGCTGCTTTCAGAGCGGCTTTCGAGTTCAGCTGTCAGAGCGGCTTTCGAGTTC	598
DB	489	GGGTCTCAGAGCGGCTTTCGAGTTCAGCTGTCAGAGCGGCTTTCGAGTTC	548
QY	599	GAGTCTATCGGCTTTCGAGTTCAGCTGTCAGAGCGGCTTTCGAGTTC	658
DB	549	CCGCGCTTTCGAGTTCAGCTGTCAGAGCGGCTTTCGAGTTC	608
QY	659	TGTGCTTTCGAGTTCAGAGCGGCTTTCGAGTTCAGAGCGGCTTTCGAGTTC	718
DB	609	TGTGCTTTCGAGTTCAGAGCGGCTTTCGAGTTCAGAGCGGCTTTCGAGTTC	668

QY 719 TCG 721
Db 669 TCG 671

RESULT 2
US-09-134-001C-537
; Sequence 537, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 537
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-537

. Query Match 4.3% Score 49.4; DB 4; Length 1065;
Best Local Similarity 45.8%; Pred. No. 7.9e-06;
Matches 213; Conservative 0; Mismatches 246; Indels 6; Gaps 1;
QY 427 GGTGCGAGCGCGGTTTGGCGAGTACGCTTTCAGGTGAGGACACACATGCTCACTG 486
Db 367 GGTTCGATGATGCAATTCGCGAGTTTACTATCTCTCTGAACAAATGTTATCATCTA 426
QY 487 CCAGACTCATTCGCCAGCATATTTGGAGCAGCTGTTGAGCCTATTTCTGTTCCGTGCAT 546
Db 427 CCAGATATATGTTTCAGCAGCAGAGAGAGTCTTGTAGAACCAACAGCCGTGCTCCAA 486
QY 547 GGTGTTGAAGCGCTGATTCAGCAGCTGCTGTCAGCGCCCTGCTTTCGAGAGAGTCTT 606
Db 487 GCAGTTAAGAGGCAATTAATTAATGCTGATGATGATGATGATGATGATGATGATGATG 546
QY 607 ATGCGCTTGCACCACTTCTGCTGCAAGGCACTCATCATGCGGCAAAATGTTGTTCC 666
Db 547 ATGCTTTGTTTACTATGTTGAGCAAAAGCTGCTGCTGCAAGTAAATATTTTCTTT 606
QY 667 GAGCGCGCTTGTATCAGAAAGACAGTTTGCAGAAAGAACTGGCGCTGAAGTGTTCATCCT 726
Db 607 GA-----CTTATCAGAAAGAACTTTAGGAAAGCTAAAGTGGTGCAGCTCAGCTG 660
QY 727 TCTACATGTAGCAGCAAAATGCTGTTCTCAGAGCTATGCTGCCGAGAGAGGATTC 786
Db 661 TATTAAGTACAGTAAAGCGATGATCAGTAAACGTTTATGAACATATGCTCAACAGCTGTA 720
QY 787 CATGACACCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846
Db 721 GATGTGATTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 847 GGACCTTCTGGAATCGCCGCTCAATGTTGGCGCTTGGGAGAGACAC 891
Db 781 CGTCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825

RESULT 3
US-09-347-803-19
; Sequence 19, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Ramodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony

APPLICANT: Orozco, Buddy
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 1610
; TYPE: DNA
; ORGANISM: Glycine max
US-09-347-803-19

. Query Match 4.3% Score 49; DB 4; Length 1610;
Best Local Similarity 48.7%; Pred. No. 1.4e-05;
Matches 133; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 443 TTGCGAGTACGTGCTTTCAGGTGAGGACACATGCTCAAGCTGCCAGACTGCTCCG 502
Db 427 TGGTATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 486
QY 503 ACATATTTGAGCAGCTGTTGAGCCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562
Db 487 TAGAGAGGAGACATGTTGATACCTTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 546
QY 563 GATTCAGCTGCTGAGAGCGCCCTGCTTTCAGAGAGGTCTATGCGCTTCCACCA 622
Db 547 ATATTGAGCAAAACAAATGTTGATGATGATGATGATGATGATGATGATGATGATGATG 606
QY 623 TTCTGCTCTGAGGCACTATGCGGCAAAATGTTGTTGCTGCTGCTGCTGCTGCTGCTG 682
Db 607 TCTGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
QY 683 GAAGACGTTTGCAGAAAGAACTGGCGCTGAAG 715
Db 667 GTTATCTGTTGCAAAATCTTGTGTCAGATG 699

RESULT 4
US-08-466-548B-1
; Sequence 1, Application US/08466548B
; Patent No. 6022712
; GENERAL INFORMATION:
; APPLICANT: sartly, aparina v
; TITLE OF INVENTION: ENHANCED YEAST EXPRESSION USING
; TITLE OF INVENTION: REGULATORY CONTROL SEQUENCES FROM YEAST SORBITOL
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,548B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/998,226
; FILING DATE: 30-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALSH, ANDREA C
; REGISTRATION NUMBER: 34988


```

? REFERENCE/DOCKET NUMBER: 95563/PS36321/US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-861-3000
? TELEFAX: 202-822-0944
? INFORMATION FOR SEQ. ID NO: 17:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1419 base pairs
? type: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
?
US-08-452-071-17

```

Query Match	3.8%	Score 43.4	DB 3	Length 1449
Best Local Similarity	66.7%	Pred. No. 0.0008		
Matches 62	Conservative 0	Mismatches 31	Indels 0	Gaps 0
QY	208	CCCTCTGTCGTCGATGATTTAGCGGAACGGTCGCGAGTGGCTCTGCTGCAC	267	
Db	281	CCCTGCTTCCTCGAGCATGACATGGTGGGAAAAGTAGTGGAGTAGGAGCATGTGTCA	340	
QY	268	AGTGTGAACCTGTCACAGAGTCCCAATTGCA	300	
Db	341	AAATTCAAGTGGGGGACACAGTTGGAATTGCA	373	

```

RESULT 11
US-09-347-803-15
; Sequence 15, Application US/09347803
; Patent No. 6274379
; GENERAL INFORMATION:
; APPLICANT: Farnodu, Layo O.
; APPLICANT: Hitz, Bill
; APPLICANT: Kinney, Tony
; APPLICANT: Orozco, Buddy
; TITLE OF INVENTION: Plant Sorbitol Biosynthetic Enzymes
; FILE REFERENCE: BB-1176
; CURRENT APPLICATION NUMBER: US/09/347,803
; CURRENT FILING DATE: 1999-07-02
; EARLIER APPLICATION NUMBER: 60/092,952
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Zea mays
; US-09-347-803-15

```

Query Match	3.7%	Score 41.8	DB 4	length 1154
Best Local Similarity	50.2%	Pred. No. 0.0033		
Matches 103	Conservative 0	Mismatches 102	Indels 0	Gaps 0
435	CGCGGATTTCGCGAGTACGTCGTTTACGGTATGAGACCAATGATCAAGCTCCAGATTC	494		
1111	1111	1111	1111	1111
501	CGGCTCCGCTGGCGAACGAGTGGTGCACCCGGCCGACCTGTCCTCAAGCTCCCGACGG	560		
1111	1111	1111	1111	1111
495	GATTCGCCGAGATTTTGGACACTGGTTAGCCTATTTTGTCTCTGGATGCTTTGA	554		
1111	1111	1111	1111	1111
561	GGTAGCCTGGAGAGGAGGCGCCATGTGCACGCGCTGACCGTGGGCTGCACGCGTGGCG	620		
1111	1111	1111	1111	1111
555	ACGCGCTAGATTTCACGCTGTGCAGACGGCCCTGTGTTCTGGAGAGAGTCTCATTCGGGCT	614		
1111	1111	1111	1111	1111
621	CCGGCGGGGGGTGGGGCCGAGACGGGGCGTCTGTGTGGGCGCCGCCCATTCGGGCT	680		
1111	1111	1111	1111	1111
615	TGCCACCATTCCTTGCTCTGCAGGC	639		
1111	1111	1111	1111	1111
681	GGTTCGCTGCTAGCGCGGAGCC	705		
1111	1111	1111	1111	1111

```
Patent No. 5451514
GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
TITLE OF INVENTION: PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/174,467
FILING DATE: 28-DEC-1993
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/874,166
FILING DATE: 27-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 95563/PS36321/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-174-467-19

Query Match          3.6%; Score 40.6; DB 1; Length 1283;
Best Local Similarity 64.2%; Pred. No. 0.006;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 207 TCCTCTGTCCTGACATGATTAGCGAAGCGTGCGAGTTGGCTCTGCTGCAC 266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 216 TCCTATGTCCTGCGCATGAAGTGTGTGAGTGTGTGAGTGTGATCAGATGTGAC 275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 267 AAGTGTGAACCTGTGTGACAGATGCGCAGTTGGANG 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 276 AAGTTCAAGTTGGAGATGTGTGCGGTGGAG 310

RESULT 13
US-08-452-071-19
Sequence 19, Application US/08452071
Patent No. 6068780
GENERAL INFORMATION:
APPLICANT: BOUDET, ALAIN M.
APPLICANT: INZE, DIRK G.
APPLICANT: SCHUCH, WOLFGANG W.
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
TITLE OF INVENTION: PLANTS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
```

```
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,071
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/874,166
FILING DATE: 27-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 95563/PS36321/US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
INFORMATION FOR SEQ. ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-452-071-19

Query Match          3.6%; Score 40.6; DB 3; Length 1283;
Best Local Similarity 64.2%; Pred. No. 0.006;
Matches 61; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 207 TCCTCTGTCCTGACATGATTAGCGAAGCGTGCGAGTTGGCTCTGCTGCAC 266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 216 TCCTATGTCCTGCGCATGAAGTGTGTGAGTGTGTGAGTGTGATCAGATGTGAC 275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 267 AAGTGTGAACCTGTGTGACAGATGCGCAGTTGGANG 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 276 AAGTTCAAGTTGGAGATGTGTGCGGTGGAG 310

RESULT 14
US-08-125-468-1/C
Sequence 1, Application US/08125468
Patent No. 5589385
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strathy, Nancy
APPLICANT: Pantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmids
TITLE OF INVENTION: useful therein
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,468
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
```


GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 09:10:26 ; Search time 171 Seconds

(without alignments)

7898.204 Million cell updates/sec

Title: US-10-020-674-1

Perfect score: 1143

Sequence: 1 atgaagagttacttataa.....cgcgcagaggttctctaa 1143

Scoring table:

IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: /publshed/Applications_NA.*
2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/PCF_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCFUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1143	100.0	1143	US-10-020-674-1	Sequence 1, Appl1
2	515.8	45.1	530	US-10-020-674-8	Sequence 8, Appl1
3	403.4	35.2	706	US-10-020-674-12	Sequence 12, Appl1
4	328	28.7	523	US-10-020-674-15	Sequence 15, Appl1
5	317.4	27.8	1158	US-10-147-003-1	Sequence 1, Appl1
6	193.2	16.9	745	US-10-147-003-7	Sequence 1, Appl1
7	103	9.0	107	US-10-020-674-11	Sequence 11, Appl1
8	83	7.3	1041	US-09-881-641-11	Sequence 11, Appl1
9	63.8	5.6	1092	US-09-815-242-7881	Sequence 7881, Ap
10	52.2	4.6	2410	US-09-925-300-386	Sequence 386, App
11	52.2	4.6	2519	US-09-969-708-281	Sequence 281, App
12	52.2	4.6	2519	US-09-969-708-281	Sequence 281, App
13	45.8	4.0	1014	US-10-025-380-153	Sequence 2189, App
14	45	3.9	513	US-10-025-380-153	Sequence 551, App
15	45	3.9	513	US-09-922-217-153	Sequence 153, App
16	45	3.9	513	US-09-833-263-153	Sequence 153, App
17	45	3.9	14333	US-08-781-986A-57	Sequence 57, Appl
18	42	3.7	2277	US-09-917-800A-1513	Sequence 1513, Ap
19	41.8	3.7	915	US-09-815-242-4651	Sequence 4651, Ap

20	41.8	3.7	1044	US-09-815-242-8368	Sequence 8368, Ap
21	39.8	3.5	1059	US-09-815-242-7877	Sequence 7877, Ap
22	39.2	3.4	1074	US-09-938-842A-1405	Sequence 1405, Ap
23	39	3.4	438	US-09-974-300-4985	Sequence 4985, Ap
24	39	3.4	1395	US-09-947-027-8	Sequence 8, Appl1
25	39	3.4	1395	US-10-091-009-8	Sequence 8, Appl1
26	39	3.4	1400	US-09-908-744-68	Sequence 68, Appl1
27	38	3.3	451	US-09-770-444-670	Sequence 670, Appl
28	37.8	3.3	453	US-09-974-300-4984	Sequence 4984, Ap
29	37.4	3.3	990	US-09-938-842A-824	Sequence 824, App
30	37	3.2	1049	US-10-123-155-358	Sequence 358, App
31	35.8	3.1	1573	US-09-971-361-8	Sequence 8, Appl1
32	35.6	3.1	395	US-09-803-719-859	Sequence 819, App
33	35.6	3.1	400	US-09-803-719-859	Sequence 819, App
34	35.6	3.1	1560	US-09-803-719-819	Sequence 2647, Ap
35	35.6	3.1	1560	US-08-781-986A-90	Sequence 90, Appl
36	35.2	3.1	451	US-09-918-995-2486	Sequence 2486, Ap
37	35.2	3.1	1015	US-09-974-300-552	Sequence 552, App
38	35.2	3.1	2061	US-10-096-960-1	Sequence 245, App
39	34.6	3.0	434	US-09-960-352-8378	Sequence 8378, Ap
40	34.6	3.0	451	US-07-728-626-2710	Sequence 2710, Ap
41	34.4	3.0	762	US-09-991-936-1554	Sequence 1554, Ap
42	34	3.0	585	US-09-833-381-1031	Sequence 1031, Ap
43	34	3.0	644	US-09-815-242-8332	Sequence 8332, Ap
44	34	3.0	1011	US-09-815-242-4441	Sequence 4441, Ap
45	34	3.0	1017	US-09-815-242-4441	Sequence 4441, Ap

ALIGNMENTS

RESULT 1
US-10-020-674-1
Sequence 1, Application US/10020674
Patent No. US20020160468A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Hiroaki
APPLICANT: Onodera, Keiko
TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
FILE REFERENCE: 06501-092001
CURRENT APPLICATION NUMBER: US/10/020,674
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: JP 2000-333363
PRIOR FILING DATE: 2000-10-31
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1143
TYPE: DNA
ORGANISM: Pichia angusta
US-10-020-674-1

Query Match	Score	DB 9	Length	DB 1143
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1143	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	1	ATGAAGGTTTACTTATAGCGTCAACGATTCCTCTCTCGAAGCGTTCTGAA	60	
DB	1	ATGAAGGTTTACTTATAGCGTCAACGATTCCTCTCTCGAAGCGTTCTGAA	60	
QY	61	CCGGAGATCAAGATCCACGATGCAAGATCAAGATCAAGATCAAGATCAAGAT	120	
DB	61	CCGGAGATCAAGATCCACGATGCAAGATCAAGATCAAGATCAAGATCAAGAT	120	
QY	121	ACGACTTGAAGATTCATATTCGAGAGTCTGTTTTCCTTAACAAGGACAC	180	
DB	121	ACGACTTGAAGATTCATATTCGAGAGTCTGTTTTCCTTAACAAGGACAC	180	
QY	181	AAGGACAGATTCGGATGCAAGATTCCTCTCTGCTGCAATGATTAACGGAAC	240	
DB	181	AAGGACAGATTCGGATGCAAGATTCCTCTCTGCTGCAATGATTAACGGAAC	240	

```

1 TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
2 FILE REFERENCE: 06501-092001
3 CURRENT APPLICATION NUMBER: US/10/020,674
4 CURRENT FILING DATE: 2002-04-19
5 PRIOR APPLICATION NUMBER: JP 2000-33363
6 PRIOR FILING DATE: 2000-10-31
7 NUMBER OF SEQ ID NOS: 17
8 SOFTWARE: FastSeq for Windows Version 4.0
9
10 SEQ ID NO: 8
11 LENGTH: 530
12 TYPE: DNA
13 ORGANISM: Pichia angusta
14
15 US-10-020-674-8
16
17 Query Match 45.1%; Score 515.8; DB 9; Length 530;
18 Match Local Similarity 98.7%; Pred. No. 4,6e-164;
19 Matches 520; Conservative 0; Mismatches 7; Indels 0; Gaps 0.
20
21 QY 274 AAACCTGTGACAGAGTCCGATTCAGCTACGCGATTCGCGACAGATCGCGCTAC 333
22 1 AAGCGGGGTGATGTCGCCACGTAAAGTACGTGCAATCTCCGACAGATCGCGCTAC 60
23 334 AAGGACACGGTCCGCCAAGACCTTGAGGGCTCTGTATGAGCTCCAGAGCGGATCTCCGAC 393
24 61 AAGGACACGGTCCGCCAAGACCTTGAGGGCTCTGTATGAGCTCCAGAGCGGATCTCCGAC 120
25 QY 394 TGCCTGCTGCGCTGAGCTTCTGCGTTTGCGTTGGGTGGTGGCCAGCGCGGTTTGGCCGAGTAC 453
26 121 TGCCTGCTGCGCTGAGCTTCTGCGTTTGCGTTGGGTGGTGGCCAGCGCGGTTTGGCCGAGTAC 180
27 QY 454 GTCGTTTACGGTGGAGACACATGATGTCAGCTGCGACAGCTGATTCGCCAGGATATTGGA 513
28 181 GTCGTTTACGGTGGAGACACATGATGTCAGCTGCGACAGCTGATTCGCCAGGATATTGGA 240
29 QY 514 GCACCTGTTGAGCCTATTCTTGTGCTCCGCGATGCTGTTGAAGCGGCTAGATTCCAGCT 573
30 241 GCACCTGTTGAGCCTATTCTTGTGCTCCGCGATGCTGTTGAAGCGGCTAGATTCCAGCT 300
31 QY 574 GGTGAGAGCGCCCTGTTCTTGTGAGAGAGTCTATCGCGCCTTGCCACCATCTTGCTCTG 633
32 301 GGTGAGAGCGCCCTGTTCTTGTGAGAGAGTCTATCGCGCCTTGCCACCATCTTGCTCTG 360
33 QY 634 CAAGCCATCATGCGGGCAAAATTGTGTCTCCGACCGCGCCTTGATCGAAGACAGTTT 693
34 361 CAAGCCATCATGCGGGCAAAATTGTGTCTCCGACCGCGCCTTGATCGAAGACAGTTT 420
35 QY 694 GGAAGGAACGTGGGCGCTGGAAGTTCGATCCTTCAATGTGACAGCAGCAATCTGTT 753
36 421 GGAAGGAACGTGGGCGCTGGAAGTTCGATCCTTCAATGTGACAGCAGCAATCTGTT 480
37 QY 754 CTCAAGGCTATGTGTCGGAGACGAGGATTCATGACGACCCCTTGA 800
38 481 CTCAAGGCTATGTGTCGGAGACGAGGATTCATGACGACCCCTTGA 527
39
40 RESULT 3
41 US-10-020-674-12
42 Sequence 12, Application US/10020674
43 Patent No. US2002016046A1
44 GENERAL INFORMATION:
45 APPLICANT: Yamamoto, Hiroaki
46 APPLICANT: Onodera, Keiko
47 APPLICANT: Tani, Yoshiki
48 TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
49 FILE REFERENCE: 06501-092001
50 CURRENT APPLICATION NUMBER: US/10/020,674
51 CURRENT FILING DATE: 2002-04-19
52 PRIOR APPLICATION NUMBER: JP 2000-33363
53 PRIOR FILING DATE: 2000-10-31
54 NUMBER OF SEQ ID NOS: 17
55 SOFTWARE: FastSeq for Windows Version 4.0
56 SEQ ID NO: 12
57 LENGTH: 706

```

TYPE: DNA
ORGANISM: Pichia angusta
US-10-020-674-12

Query Match 35.28; Score 402.4; DB 9; Length 706;
Best Local Similarity 99.8%; Pred. No. 1.4e-125;
Matches 403; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 740 ACGCAATGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 799
DB 3 ATGCAATGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 62
OY 800 ACTGCTGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 859
DB 63 ACTGCTGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 122
OY 860 TGGCCGCTGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 919
DB 123 TGGCCGCTGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 182
OY 920 CTACAGGAGAAATACGCTACCGGCTCATGCTACACCGCTCAAGGAGGATTCAGGAG 979
DB 183 CTACAGGAGAAATACGCTACCGGCTCATGCTACACCGCTCAAGGAGGATTCAGGAG 242
OY 980 TTGTCAAGGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 1039
DB 243 TTGTCAAGGCTGTTCTCAAGGCTATGGTGGGAGAGAAAGGAGGATTCATGACGCTTCG 302
OY 1040 GCAAGTCCACCTTAAGGAGAGGAGTGGAGAGGAGGATTCATGACGCTTCGAGGAG 1099
DB 303 GCAAGTCCACCTTAAGGAGAGGAGTGGAGAGGAGGATTCATGACGCTTCGAGGAG 362
OY 1100 AGACAAATGTCAGATCTGTTGAGAGCCGCAAGGATTCATGACGCTTCG 1143
DB 363 AGACAAATGTCAGATCTGTTGAGAGCCGCAAGGATTCATGACGCTTCG 406

RESULT 4

US-10-020-674-15

Sequence 15, Application US/10020674
Patent No. US20020160468A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Hiroaki
APPLICANT: Onodera, Keiko
APPLICANT: Tanai, Yoshiaki
TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
FILE REFERENCE: 06501-092001
CURRENT APPLICATION NUMBER: US/10/020,674
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: JP 2000-333363
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 15
LENGTH: 523
TYPE: DNA
ORGANISM: Pichia angusta
US-10-020-674-15

Query Match 28.7%; Score 328; DB 9; Length 523;
Best Local Similarity 100.0%; Pred. No. 1.9e-100;
Matches 328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGAAGGTTACTTATTTAGGATCAAGATATTCGCTACTCCGAACGGTTCTGAA 60
DB 192 ATGAAGGTTACTTATTTAGGATCAAGATATTCGCTACTCCGAACGGTTCTGAA 251
OY 61 CCGAGATCAAGATCCCAAGATGTCAGATCAAGGATTCGCTATTTGGGAAATGAGG 120
DB 252 CCGAGATCAAGATCCCAAGATGTCAGATCAAGGATTCGCTATTTGGGAAATGAGG 311
OY 121 ACGGATGGAAGATTCATATTTGAGAGCTCTGTTTTCCTTAACAAGGAGCACC 180

DB 312 ACGGATGGAAGATTCATATTTGAGAGCTCTGTTTTCCTTAACAAGGAGCACC 371
OY 181 AAGCAAGATTTCCGGATACGAACTTCTCTGCTCCGACATGAAATTTAGCGGAGC 240
DB 372 AAGCAAGATTTCCGGATACGAACTTCTCTGCTCCGACATGAAATTTAGCGGAGC 431
OY 241 GTGCTGAGGTTGGCTCTGCTGTCACAGTGAACCTGTCAGAGATGTCAGTTGAA 300
DB 432 GTGCTGAGGTTGGCTCTGCTGTCACAGTGAACCTGTCAGAGATGTCAGTTGAA 491
OY 301 GGTAGTGCATTTGCTCCGACAGATCCG 328
DB 492 GGTAGTGCATTTGCTCCGACAGATCCG 519

RESULT 5

US-10-147-003-1

Sequence 1, Application US/10147003
Publication No. US20030032153A1
GENERAL INFORMATION:
APPLICANT: Yamamoto, Hiroaki
APPLICANT: Kimoto, No. US20030032153A1Hiro
TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE.
TITLE OF INVENTION: METHODS FOR PRODUCING SAME, AND METHODS FOR PRODUCING
FILE REFERENCE: 06501-105001
CURRENT APPLICATION NUMBER: US/10/147,003
CURRENT FILING DATE: 2002-05-16
PRIOR APPLICATION NUMBER: JP 2001-159647
PRIOR FILING DATE: 2001-05-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1158
TYPE: DNA
ORGANISM: Kluyveromyces fragilis
US-10-147-003-1

Query Match 27.8%; Score 317.4; DB 9; Length 1158;
Best Local Similarity 56.3%; Pred. No. 1.2e-96;
Matches 638; Conservative 0; Mismatches 486; Indels 9; Gaps 2;

OY 1 ATGAAGGTTACTTATTTAGGATCAAGATATTCGCTACTCCGAACGGTTCTGAA 60
DB 1 ATGAAGGTTACTTATTTAGGATCAAGATATTCGCTACTCCGAACGGTTCTGAA 60
OY 61 CCGAGATCAAGATCCCAAGATGTCAGATCAAGGATTCGCTATTTGGGAAATGAGG 120
DB 61 CCGAGATCAAGATCCCAAGATGTCAGATCAAGGATTCGCTATTTGGGAAATGAGG 120
OY 121 ACGGATGGAAGATTCATATTTGAGAGCTCTGTTTTCCTTAACAAGGAGCACC 180
DB 121 ACGGATGGAAGATTCATATTTGAGAGCTCTGTTTTCCTTAACAAGGAGCACC 180
OY 181 AAGCAAGATTTCCGGATACGAACTTCTCTGCTCCGACATGAAATTTAGCGGAGC 240
DB 175 AAGCAAGATTTCCGGATACGAACTTCTCTGCTCCGACATGAAATTTAGCGGAGC 240
OY 241 GTGCTGAGGTTGGCTCTGCTGTCACAGTGAACCTGTCAGAGATGTCAGTTGAA 300
DB 241 GTGCTGAGGTTGGCTCTGCTGTCACAGTGAACCTGTCAGAGATGTCAGTTGAA 300
OY 235 GTATCAAAAATGAGGAGGATTCACAACTGTTGGCCATTTGGGTTTCATGGGT 414
DB 235 GTATCAAAAATGAGGAGGATTCACAACTGTTGGCCATTTGGGTTTCATGGGT 414
OY 421 TTGGTGGTGGCCAGGCGGTTTGGCCAGATGAGTGTGTTAGCGTGAAGACATGATC 480
DB 415 TTGGTGGTGGCCAGGCGGTTTGGCCAGATGAGTGTGTTAGCGTGAAGACATGATC 474

Query Match	16.9%;	Score 193.2;	DB 9;	Length 745;
Best Local Similarity	55.4%;	Pred. No. 1e-54;		
Matches 397;	Conservative	0;	Mismatches 313;	Indels 6;
			Gaps	1

```

RESULT 7
US-10-020-674-11
; Sequence 11, Application us/10020674
; Patent No. US20020160468A1
; GENERAL INFORMATION:
; APPLICANT: Yamamoto, Hiroaki
; APPLICANT: Onodera, Keiko
; APPLICANT: Tani, Yoshiaki
; TITLE OF INVENTION: NOVEL (R)-2,3-BUTANEDIOL DEHYDROGENASE
; FILE REFERENCE: 06501-092001
; CURRENT APPLICATION NUMBER: US/10/020,674
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: JP 2000-333363
; PRIOR FILING DATE: 2000-10-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 107
; TYPE: DNA
; ORGANISM: Pichia angusta
US-10-020-674-11

Query Match      9.0%; Score 103; DB 9; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;

```



```
QY 852 TTCTGGAATCGCCGTCATGTCGCGCTTGGGAGACCAACCAATTGATTCACCAAT 911
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 822 GGCCGGCAATATGCTGCTTGTGGGAAATCT---TCGAGAGGCCCAAGTAATTTCTT 878
QY 912 GTCTGCACTTCCAGGAAATAGCCTACCGCTCCATGCTTACACGTCGAAGACTT 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 879 CGAGCTGCTTCCAGGAGAGAGATGCTCGGGGCCCTCGCTTACAAAGCGGAGTTCCG 938
QY 972 CCAGGAAGTTTCAGAGCCCTTGGAGAGATGCTCATATCTTTGGACAAAGCCGCAAGAT 1031
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 939 CGAGGATGCGCTTATAGCCAGCGCGCCCT-----CGACATGCGACCGCT 986
QY 1032 GATTACAGCAAGTCCACTTAAGACGAGTCGAGAGGCTTTAAACGCTGATGCA 1091
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 987 GGTGACCGGTCGCGCTGGAAGAAATCTGCGAGCGCGGCTTTCGAAGAAGTTGTGAA 1046
QY 1092 GCACAAGAGACATGTCTAGATCTGTGTGACGCCGCAAGC 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1047 CAACAAGAGACACACGTGAAGATCTGTCTGCGCGGGCG 1087
```

```
RESULT 10
US-09-925-300-386
; Sequence 386, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 386
; LENGTH: 2410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2167)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-386
```

```
Query Match 4.6%; Score 52.2; DB 10; Length 2410;
Best Local Similarity 51.0%; Pred. No. 1.2e-06;
Matches 123; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 480 CAACCTGCGACACTGCATTCGCCGAGCATTTGAGCACTGGTGTATTTCTGTGC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 CAACCTTCTTCAATGTCATCTTGAAGAGAGCGCCCTGATGACGACCTTTGTGGG 480
QY 540 CTGCATGCTGTGACGCGCTAGATTCACGCTGTGTGACGCGCCCTGTCTTGGAGG 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 GATCATCTCCCTGAGAGAGCGGAGGATTAACCTGGGACACAAAGTCTGTGTGAGG 540
QY 600 AGGTCTATCGGCGCTTGCACCACTTCTGTCTGCAAGGCGCATATCGGGCAAAATGT 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 TGGGCAATCGGAGATGTCACCTTCTGTCTGTGCGCAAGCAATGGAGAGCTCAAGTGT 600
QY 660 GTGTTCGAGCGCGCTGATCAGAGACAGTTCGAAGAGACTGGCGCTGAAGTGT 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 GGTGACTGATCTGTCTTACCGGATTTGTCCAAAGCCAAAGAGATGGGCTGATTTAGT 660
QY 720 C 720
    |
Db 661 C 661
```

```
RESULT 11
US-09-969-708-281
; Sequence 281, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-70
; CURRENT APPLICATION NUMBER: US/09/969,708
; PRIOR FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: US/60/237,606
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,608
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: US/60/237,425
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 281
; LENGTH: 2519
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-969-708-281
```

```
Query Match 4.6%; Score 52.2; DB 10; Length 2519;
Best Local Similarity 51.0%; Pred. No. 1.2e-06;
Matches 123; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 480 CAACCTGCGACACTGCATTCGCCGAGCATTTGAGCACTGGTGTATTTCTGTGC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 560 CAAGCTTCTTCAATGTCATCTTGAAGAGAGCGCCCTGATGACGACCTTTGTGGG 619
QY 540 CTGCATGCTGTGACGCGCTAGATTCACGCTGTGTGACGCGCCCTGTCTTGGAGG 599
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 620 GATCATGCTCAGAGAGAGGAGGAGTACCTGGGACACAAAGTCTGTGTGAGG 679
QY 600 AGGTCTATCGGCGCTTGCACCACTTCTGTCTGCAAGGCGCATATCGGGCAAAATGT 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 680 TGGGCAATCGGAGATGTCACCTTCTGTCTGTGCGCAAGCAATGGAGGACGCTCAAGTGT 739
QY 660 GTGTTCGAGCGCGCTGATCAGAGACAGTTCGAAGAGACTGGCGCTGAAGTGT 719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 GGTGACTGATCTGTCTTACCGGATTTGTCCAAAGCCAAAGAGATGGGCTGATTTAGT 799
QY 720 C 720
    |
Db 800 C 800
```

```
RESULT 12
US-09-880-107-2189
; Sequence 2189, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2189
; LENGTH: 2519
; TYPE: DNA
```

GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Lodes, Michael J.
APPLICANT: Secrist, Heather
APPLICANT: Benson, Darin R.
APPLICANT: Meagher, Madeline Joy
APPLICANT: Stolk, John A.
APPLICANT: Wang, Tongtong

```

; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aljun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FASTSEQ FOR Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-922-217-153

```

```

Query Match          3.9%; Score 45; DB 10; Length 513;
Best Local Similarity 51.2%; Pred. No. 0.00013;
Matches 105; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

```

QY 480 CAGCTGCCAGACTCGATTCGCGAGATTTGGAGCAGCTGTTGAGCTATTCTGTGC 539
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
DB 304 CAGCTTCTGACATGTCACTTTGAGAGAGCGCCCTGATGAGCACTTCTGTGGG 363
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
QY 540 CTGGCATGCTGTGACGCGCTAGATTCCAGCTGCTGAGAGCGCCCTGTTCTTGAGG 599
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
DB 364 GATCCATGCTGCGAGAGAGCGAGTTACCTGCGACACAGGTCCTGTGTGTGAGG 423
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
QY 600 AGTCCCTATGGCCCTTCCACCATTTCTGCTGCAAGGCCATCATGCGGCAAAATTGT 659
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
DB 424 TGGGCCAATGCGGATTTGCTGCTGCGCAAAAGCAATGGAGCAGCTCAAGTAGT 483
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
QY 660 GTTCCGAGCGGCTTGTATCAGA 684
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |
DB 484 GGTGACTGATCTGTCTGTAACCGA 508
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: May 2, 2003, 12:08:43
Job time : 177 secs